



UnB

**UNIVERSIDADE DE BRASÍLIA (UNB)
FACULDADE DE CIÊNCIA DA INFORMAÇÃO (FCI)**

BIG DATA: FUNDAMENTOS, METODOLOGIAS E TECNOLOGIAS

Glenda Rany Máximo de Souza

Orientador: Dr. Ailton Luiz Gonçalves Feitosa

**Brasília
2017**

GLENDA RANY MÁXIMO DE SOUZA

BIG DATA: FUNDAMENTOS, METODOLOGIAS E TECNOLOGIAS

Monografia apresentada à banca examinadora como requisito parcial para a conclusão do curso de Biblioteconomia da Faculdade de Ciência da Informação, Universidade de Brasília.

Orientador: Dr. Ailton Luiz Gonçalves Feitosa

Brasília

2017

S729 Souza, Glenda Rany Máximo de.

Big Data: fundamentos, metodologias e tecnologias /
Glenda Rany Máximo de Souza. – Brasília, 2017. – 63f.
Monografia (Bacharelado em Biblioteconomia) –
Universidade de Brasília, Faculdade de Ciência da
Informação, 2017. – Orientação: Ailton Luiz Gonçalves
Feitosa

1. Big Data. 2. Sistemas de Informação Gerencial. 3.
Dilúvio informacional. 4. Biblioteca I. Título.



Título: Big Data: fundamentos, metodologias e tecnologias

Aluno: Glenda Rany Máximo de Souza

Monografia apresentada à Faculdade de Ciência da Informação da Universidade de Brasília, como parte dos requisitos para obtenção do grau de Bacharel em Biblioteconomia.

Brasília, 30 de janeiro de 2017.

Ailton Luiz Gonçalves Feitosa – Orientador
Professor da Faculdade de Ciência da Informação (UnB)
Doutor em Ciência da Informação

Maria da Conceição Lima Afonso – Membro externo
Analista no Serviço Nacional de Aprendizagem Industrial (SENAI/DN)
Mestre em Ciência da Informação

José Marcelo Schiessi – Membro externo
Analista na Caixa Econômica Federal (CEF)
Doutor em Ciência da Informação

RESUMO

O estudo analisa os conceitos e definições envolvidos no modelo Big Data. Contextualiza a evolução histórica do Big Data e descrevem os fundamentos, metodologias e tecnologias utilizadas pela tecnologia Big Data. Apresenta o Big Data como recurso que tem auxiliado diversos serviços informacionais. Com a grande quantidade de volume informacional existente há uma real necessidade de tratar os dados para se transformar em informação útil. A partir da coleta, do processamento e do armazenamento de dados às análises realizadas por todos os centros e serviços de documentação têm-se tornado objeto de estudo e aperfeiçoamento. Conclui que há aplicações possíveis em diferentes áreas, com potencial aplicação em bibliotecas. Adverte para fatores como custo de gerenciamento do BD e perda de privacidade das pessoas em geral.

Palavras-chave: Big Data; Sistemas de Informações Gerenciais; Volume de dados; Tratamento de dados.

ABSTRACT

The study analyzes the concepts and definitions involved in the Big Data model. Contextualizes the historical evolution of Big Data and describes the fundamentals, methodologies and technologies used by Big Data technology. It presents Big Data as a resource that has aided several information services. With the large amount of informational volume there is a real need to treat the data to become useful information. At from the collect, of the processing and storage of data, the analyzes realized by all the centers and documentation services have become the object of study and perfectioning. It concludes that there are possible applications in different areas, with potential application in libraries. Warns about factors such as cost of managing BD and loss of privacy of people in general.

Keywords: Big Data; Management Information Systems; Volume of data; Treatment of data.

LISTA DE FIGURAS

Figura 1 – Diagrama geral do sistema de comunicação	19
Figura 2 – Características do Big Data – 3V's.....	22
Figura 3 – Os 5 V's.....	25
Figura 4 – Fluxograma SaaS.....	31
Figura 5 - Como funciona o Hadoop	36
Figura 6 - Como funciona MapReduce.....	37
Figura 7- Cloud Computing	38
Figura 8 - Aplicações do Big Data nas empresas.....	43
Figura 9 - Características nas Bibliotecas	50

LISTA DE QUADROS

Quadro 1 - Sistema de Informações Gerenciais.....	27
Quadro 2 - Visão geral das tecnologias de Big Data.....	35
Quadro 3 – Analytics	40

LISTA DE SIGLAS E ABREVIATURAS

BD	Big Data
BI	Business Intelligence
DM	Data Mining
DW	Data Warehouse
ERP	Enterprise Resources Planning
HDFS	Hadoop Distributed File System
HTML	Hyper Text Markup Language
HTTP	Hypertext Transfer Protocol
IBM	International Business Machines
IoT	<i>Internet</i> of Things
MRP	Material Requirement Planning
MRP II	Manufacturing Resources Planning
RFID	Radio Frequency Identification
SaaS	Software as a Service
SGBDS	Sistemas de Gerenciamento em Banco de Dados
SIG	Sistemas de Informação Gerencial
SOA	Service-Oriented Architecture
URI/ URL	Uniform Resource Identifier
TI	Tecnologia da Informação
2 V's	Variabilidade e Veracidade
3 V's	Volume, Variedade e Velocidade
WSN	Wireless Sensor Networks
WWW	World Wide Web Foundation

SUMÁRIO

1	INTRODUÇÃO	12
2	JUSTIFICATIVA	14
3	OBJETIVOS	15
3.1	OBJETIVO GERAL.....	15
3.1.1	Objetivos Específicos	15
4	METODOLOGIA.....	16
5	BIG DATA	17
5.1	HISTÓRICO	17
5.1.1	A era analógica do processamento de dados.....	17
5.1.2	A era de tramitação de processamento de dados	19
5.1.3	A era de consolidação de processamento de dados	20
5.2	DEFINIÇÕES	21
5.2.1	Volume.....	22
5.2.2	Variedade	23
5.2.3	Velocidade.....	24
5.2.4	Os 2V's a mais	24
5.2.5	Variabilidade	25
5.2.6	Veracidade	26
5.3	SISTEMA DE INFORMAÇÃO GERENCIAL	26
5.3.1	ERP (Enterprise Resource Planning).....	27
5.3.2	BI (Business Intelligence)	28
5.3.3	SaaS (Software as a Service).....	30
5.3.4	SOA (Service-Oriented Architecture).....	31
5.3.5	Internet of Things (IoT).....	32
5.4	TECNOLOGIAS UTILIZADAS.....	34
5.4.1	Hadoop	35
5.4.2	MapReduce.....	36
5.4.3	Cloud Computing.....	37
5.4.4	Analytics.....	40
5.5	APLICAÇÕES DO BIG DATA.....	41
5.5.1	IBM	43
5.5.2	Facebook.....	45
5.5.3	Amazon.....	46
5.5.4	Google	47
5.6	APLICAÇÕES EM BIBLIOTECAS.....	49
5.6.1	Fonte de dados	50
5.6.2	Arquitetura e Sistemas de dados	52
5.6.3	Infraestruturas e Suporte tecnológicos.....	53
5.6.4	Capacitação.....	53

5.6.5	Estrutura Organizacional	55
5.6.6	A biblioteca, dados e informação.....	56
6	CONSIDERAÇÕES FINAIS	58
	REFERÊNCIAS.....	60

1 INTRODUÇÃO

A informação é considerada a moeda que tem movido o mundo nessa nova era tecnológica. O poder que se pode obter administrando de forma favorável às informações é o novo recurso para habilitar o aperfeiçoamento de um serviço. A grande movimentação de dados tem exposto a defasagem que ocorre em seu processo de armazenamento em massa. Assim, como a necessidade de otimizar o processamento de dados e as tecnologias desenvolvidas.

Antes da Internet, o método utilizado para análise e propagação da informação era longo, demorado e muitas vezes incerto. Com o surgimento da Internet, as formas de lidar com a informação foram aprimoradas. Com a melhoria de ferramentas, a necessidade de conhecimento foi crescendo e a demanda foi se tornando maior. O volume de informação expandiu-se e trouxe novas questões aos sistemas.

Essa intensificação do fluxo de dados é nomeada de Big Data (BD) e aparece pela primeira vez em um artigo publicado por dois pesquisadores da Nacional Aeronautics and Space Administration (NASA) em 1997. Michael Cox e David Ellsworth trazem o que seria a ideia do Big Data, nesse primeiro olhar, mencionando uma exacerbação de dados. De fato, chama-se como “O problema do Big Data”; que trata a falta de espaço de memória principal, influenciando o processamento de dados e, como o seu armazenamento seria complexo. (COX; ELLSWORTH, 1997).

O Big Data é um modelo que tem se tornado o diferencial quando se pensa em tratamento de dados. Com a capacidade de gerenciar rapidamente grandes volumes de dados de diferentes origens, ele é o referencial para onde estão se encaminhando as prioridades de conhecimento. E é a partir dessa curiosidade de como se dará o mundo depois das transformações que acontecerão, que partirá o estudo de como será suas aplicações, definindo suas características históricas, determinando o conceito do termo e sua relação com as empresas.

As instituições movimentam-se para aprimorar os procedimentos cabíveis para se adequar a uma arquitetura de informação diferenciada. Nesse processo, não só as tecnologias auxiliam as organizações como também determinam o nível de adaptação com os diferentes estilos e formatos de

dados analisáveis. Tornando assim, a nova preocupação da sociedade de informação responsável em lidar com toda essa nova estrutura informacional.

2 JUSTIFICATIVA

O presente trabalho esclareceu a importância de entender como o Big Data pode ser primordial para a construção de tecnologias que favoreçam o trabalho dos profissionais da informação. Ao compreender que as demandas informacionais mudaram, assumiu-se a relevância de como a modernização é essencial para o campo da ciência da informação. E também, elucidou os processos básicos que foram precisos para se adequar a esses progressos.

Apresentou ainda que a necessidade de auxiliar os profissionais da informação em se adaptar aos diversificados meios tecnológicos de processamento de dados é real. Levantou conceitos históricos e estruturais que demonstraram a abrangência do Big Data. Além disso, esclareceu como se deram os fatores que foram influenciadores em adequar o Big Data (BD) nas empresas e assim comparou também em como aplicar nas bibliotecas. Demonstrando assim a relevância de trazer o BD a realidade dos centros de documentos.

Com a obrigatoriedade de sempre se atualizar e adaptar, os especialistas da disseminação da informação, são o público alvo que o trabalho alcançou. Como as tecnologias são as responsáveis em estimular tantas competências, a área de Tecnologia da Informação (TI), é envolvida em todo e qualquer tipo de trabalho científico realizado nas instituições. Por motivo desse estreitamento entre as áreas de TI e as áreas de Ciência da Informação, os profissionais encarregados das esferas de organização informacional, como os bibliotecários são os interessados nas aplicações que são manifestadas e atualizadas diariamente que, agilizam e facilitam parte do seu trabalho.

Em geral o caráter histórico, conceitual, estrutural e aplicável que permeiam a integridade desse tema teve embasamento em Davenport (2014). Ele teoriza e exemplifica como o Big Data tem potencialidade em transformar qualquer sistema de informação. Contextualizando cada um dos aspectos que a tecnologia abrange dá abertura para a otimização dos serviços e produtos que possam ser oferecidas relacionadas à modernização.

3 OBJETIVOS

3.1 OBJETIVO GERAL

- Identificar as características do Big Data aplicáveis nas bibliotecas.

3.1.1 Objetivos Específicos

- Contextualizar a evolução histórica do Big Data.
- Sistematizar as características dos conceitos de Big Data.
- Explorar as tecnologias que contribuíram para o Big Data.
- Explorar as tecnologias desenvolvidas a partir do Big Data.
- Identificar aplicações Big Data a bibliotecas

4 METODOLOGIA

O presente trabalho se organiza em uma pesquisa com nível de aprofundamento exploratório, cujo intuito é abordar a tecnologia Big Data. Onde em um primeiro momento se detém em contextualizar a evolução histórica de seu advento assim como, a sua definição propriamente. E em segundo, se estende em investigar a sua aplicação e as tecnologias envolvidas. (GIL, 2002).

Utiliza-se, neste trabalho, a técnica de levantamento bibliográfico onde são explanadas as questões históricas, definições, aplicações e ferramentas utilizadas pelo modelo Big Data. Como método de sistematização dos conhecimentos, adotou-se o monográfico. (GIL, 2002).

5 BIG DATA

5.1 HISTÓRICO

Historicamente a sociedade vem se transformando por causa dos ciclos informacionais. Conforme as demandas foram se modificando, os processos de tratamento de informações tiveram que se adequar. Partindo da ideia de que as épocas são delimitadas por meio das similaridades destacadas pela a evolução tecnológica que cada tempo possui e que, cada uma é modificada depois que um novo método a revoluciona, foi possível determinarem os limites de cada momento na história e classificá-las em três períodos: a era **analógica**, de **tramitação** e de **consolidação** do processamento de dados.

Cada era possui um início e um ápice para determinar sua duração. A era analógica foi determinada pelos primórdios dos métodos de coleta de informação, quando a civilização começou a demonstrar indícios de preocupação com a sua apuração até a inserção dos computadores nas empresas. A era de tramitação começa na disseminação dos computadores pessoais até ao aumento de velocidade de produção e volume de informação. E por fim a era de consolidação, na qual o tratamento de informação começa a se dissipar em desenvolvimento de várias aplicações.

5.1.1 A era analógica do processamento de dados

O ábaco é uma invenção datada e (re) criada por diversos povos desde 4.000 a.C. Foi a primeira ferramenta precisa de contabilização e processamento de dados. Sendo um instrumento para época como uma grande solução, ele com tempo foram trazendo a necessidade de processar dados de outros modos. (GODOI, 2015).

A preocupação em poder apurar grandes quantidades de dados faz o ano de 1880 nos Estados Unidos, ser o ano da nova ordem de processamento. Havia a necessidade de fazer um censo no qual seria possível averiguar informações sobre a população. Ao concluir somente a parte da coleta de dados, percebeu-se que prolongaria por mais oito anos, para apenas **tabular/tabelar** esses dados. E foi previsto que o censo de 1890 se alongaria por mais 10 anos para ser finalizado. (MAYER - SCHÖNBERGER; CUKIER, 2013).

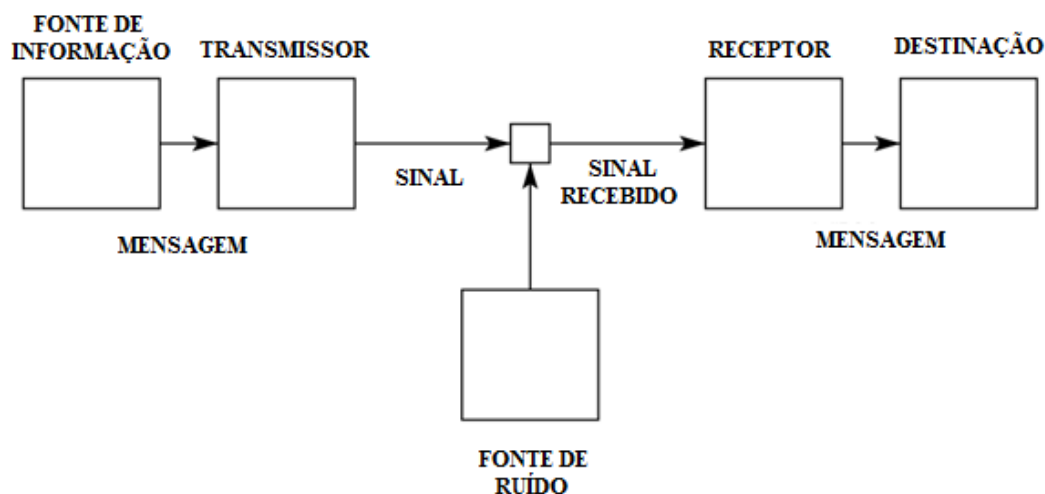
Entretanto em 1896, com o Herman Hollerith e com seus cartões perfurados, inventa-se a tabela de Hollerith, a *Tabulating Machine*, que resultou em um método de tratamento de dados e acelerou razoavelmente os meios de analisá-los. Essa *Máquina Tabuladora* possuía o princípio básico de como seria/era a análise de dados em uma máquina. (IBM, 2016; ALMEIDA, 1997).

O método acelerou a produção de informações e percebeu-se que o volume produzido ultrapassaria a rapidez com que seria digerida pelos cartões perfurados. Fora a sua aplicabilidade em diversas áreas que aumentou ainda mais a produção de informação. Mesmo assim, ainda com essa produção, esse meio não conseguiria interpretar tantos fluxos de informações a tempo. Então, o que não conseguiria ser dimensionados pelos os cartões perfurados, tornou-se a base de outra tecnologia, a tecnologia de máquinas. (ALMEIDA, 1997).

Em 1944, com a explosão da informação, as bibliotecas, centros de documentação e arquivos tiveram que se adaptar ao moderno parâmetro de armazenamento que começa a originar das máquinas. Com isso, toda essa expansão de informação, tornou o seu gerenciamento mais trabalhoso devido à quantidade informacional, entretanto os métodos de tratamento estavam começando a despontar. (WINSHUTTLE, 2016).

Em 1948, Claude Shannon produz um artigo chamado *A mathematical theory of communication*, que até hoje serve de base da ideia de processamento de dados moderno. De acordo com a Figura 1 retirado do artigo, demonstra uma estrutura que determina os requisitos mínimos de dados para transmitir informações através de canais ruidosos, e como o canal consegue fazer conexão entre a fonte e sem causar alteração neles. (SHANNON, 1948; WINSHUTTLE, 2016).

Figura 1 – Diagrama geral do sistema de comunicação



Fonte: (SHANNON, 1948)

Ao ajustar como funcionaria a forma de processamento com que os dados traduziriam as mensagens, a sua armazenagem por outro lado, causariam outra preocupação. Os meios e locais de armazenamento e a forma como seriam alocados, trouxeram esse receio. Em 1960, setores não só da ciência como de empresas começam a pensar em como criar máquinas que organizariam todo esse conjunto de questões sobre a guarda de dados. (WINSHUTTLE, 2016).

Em 1970, cria-se o conceito de um banco de dados relacional, publicado por um pesquisador da IBM, onde ele pede por informações armazenadas dentro de um computador para ser organizados de modo fácil para interpretar tabelas de forma que os usuários não técnicos pudessem gerenciar e acessar grandes quantidades de dados. Os computadores pessoais já começavam a ganhar popularidade, e as empresas já estavam desenvolvendo ferramentas que as auxiliariam em organizar e agendar suas informações. (WINSHUTTLE, 2016).

5.1.2 A era de tramitação de processamento de dados

Em 1980, as empresas aperfeiçoam a forma de processar seus dados até transcender e começarem a produzir sistemas que se adequam em auxiliar, criar, organizar e agendar suas informações. Esses sistemas padronizaram os procedimentos e foram sendo aprimorados em um novo produto que foi

denominado *Material Requirement Planning (MRP)*. Inicia-se um aspecto diferente de tratar esses dados, e então um conceito novo começa a ser difundido pelos Estados Unidos da América (EUA), sendo apresentado como o novo conceito passou a ser difundido pelos EUA, sendo apresentado como uma possibilidade às práticas convencionais de gerenciamento de estoques na produção industrial. (LAURINDO, 2000; WINSHUTTLE, 2016).

Pouco tempo depois em 1985, o MRP evolui para o *Manufacturing Resources Planning (MRP II)*, pois foi preciso detalhar os subprocessos para verificação da capacidade de produção. E assim, chegando aos anos 90, com o modelo *Enterprise Resources Planning (ERP)* uma nova geração de sistemas que tem sua abrangência expandida para outras áreas e que consegue enfim, trazer em um programa só as análises necessárias a serem consideradas por toda empresa. (LAURINDO, 2000).

Em meados dos anos 90 o computador já estava consolidado nas operações cotidianas tanto de empresas quanto de pessoas e o termo Business Intelligence (BI), expande-se e se desenvolve nas empresas responsáveis por administrar especificamente banco de dados relacional e variáveis da própria organização. (WINSHUTTLE, 2016).

Em outro plano, Tim Berners-Lee criava uma proposta que unificava um canal direto de informação, a Web. Em março de 1989, Berners-Lee expôs sua visão do que seria Web publicando o artigo "Gestão de Informação: Uma Proposta". Nesse trabalho, ele descreve as 3 tecnologias que fundamentam a Web, que são as HyperText Markup Language (HTML), Uniform Resource Identifier (URI/ URL) e Hypertext Transfer Protocol (HTTP). (WORLD WIDE WEB FOUNDATION, 2016).

E então depois da materialização da Web, em 1997 o termo Big Data (BD) aparece pela primeira vez em um artigo, publicado por dois pesquisadores Michael Cox e David Ellsworth. (WINSHUTTLE, 2016).

5.1.3 A era de consolidação de processamento de dados

Logo após de lançado o termo Big Data, sua conceituação transformou-se em explicação para qualquer evento relacionado à tecnologia de *datificação* (transformar tudo em dados). E então, tecnologias que despontaram a partir

dos anos 2000 foram agregadas com a capacidade de manipulação de dados. (DAVENPORT, 2014).

Em 1999, a expressão “*Internet of Things*” (*Internet das coisas*), surge e apresenta como serão controlados os sistemas de informações de dispositivos por meio de outros dispositivos conectados via *Internet*. Em 2001, o SaaS (Software as a Service) e sua metodologia facilmente aplicada a Cloud Computing surgem dando abertura para outras formas de *software* inteligentes. Delimita-se as dimensões do Big Data, apresentando os 3V’s de Doug Laney. Em 2006, cria-se o Hadoop que consolida como um novo software que é aplicado exclusivamente ao gerenciamento e armazenamento de dados. E assim, consagra-se a nova remessa de tecnologias fundamentada no novo estilo de processamento de dados adaptadas à realidade da nova sociedade de informação. (WINSHUTTLE, 2016; TAURION, 2015).

5.2 DEFINIÇÕES

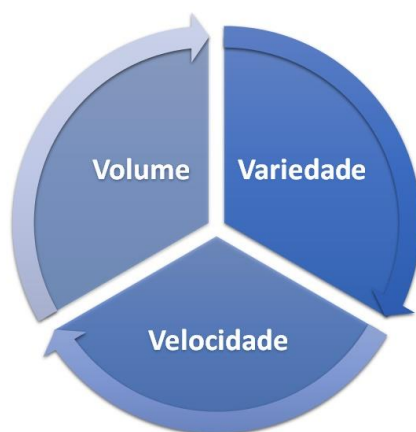
Como mencionado anteriormente, o Big Data surge como um problema identificado por dois pesquisadores da NASA. Michael Cox e David Ellsworth. Em 1997, esses autores publicaram um artigo abordando sobre como o volume de dados acessados ultrapassava a quantidade da memória interna do computador que a processava. Em outras palavras, o volume de informação era tão grande que a memória principal não conseguia processar rapidamente inúmeras informações, porque não tinha para onde expandir. Trata-se também, sobre como a demanda de informação aumentava e sobre como as pessoas estavam produzindo e divulgando muito mais conhecimento. (COX; ELLSWORTH, 1997).

Com o aumento da quantidade de informação, cientistas e pesquisadores da área de tecnologia se aprofundam e começam a vincular novas pesquisas acerca do que seria BD. Até então, o Big Data era atribuído a tudo aquilo que não conseguiam explicar e definir dentro das novas tecnologias. (COX; ELLSWORTH, 1997).

Ao desenvolver a arquitetura do Big Data, foram se contextualizando as primeiras definições de BD. Uma das mais consagradas definições é sobre a grande massa de dados brutos fluindo pelos sistemas. Ou um grupo de

tecnologias lidando com as várias informações das organizações em um fluxo contínuo. Entre outras características, foram surgindo as questões das variações de dados, e como não só o volume era uma característica fixa e única do BD. (DAVENPORT, 2014; MAYER - SCHÖNBERGER; CUKIER, 2013).

Figura 1 – Características do Big Data – 3V's



Fonte: (IVANOV; KORFIATIS; ZICARI, 2013)

Destacam-se também ideias sobre a mutabilidade de dados e a duração do tempo de resposta dos dados dentro do sistema. Percebe-se que não apenas, a característica de volume o define, mas outros dois fatores influenciam a sua estrutura. É o que se observa na figura 2. São determinadas, então, mais duas extensões do BD que, conjuntamente, formam os três segmentos chamados de 3V's: **Volume, Variedade e Velocidade**. (LANEY, 2001; DAVENPORT, 2014).

5.2.1 Volume

A característica **Volume**, como dito anteriormente, foi um dos primeiros atributos. A partir do número de dados que foram aparecendo demasiadamente dentro dos sistemas, foram se desenvolvendo métodos para auxiliar na administração do fluxo de dados. Entre esses métodos, a ideia do que era Big Data ganhou corpo, e tornou-se uma tecnologia. (MAYER - SCHÖNBERGER; CUKIER, 2013).

A tecnologia BD quanto ao atributo volume, define-o em dados brutos em grande fluxo, significando que o tamanho do conjunto era apenas dados. O volume é exemplificado diariamente pela alta quantidade de informação acessada, trocada e reutilizada dentro de redes e/ou sistemas. (SCUDERE, 2015; MAYER - SCHÖNBERGER; CUKIER, 2013).

A quantificação da informação advinda de vários canais de informações com milhares de usuários, sendo alimentada diariamente de toda e qualquer movimentação no ambiente digital é considerada volume. Assim como, o montante de dados aglomerados em diversos servidores e que se correspondem, é a essência de como se organiza a característica de volume. (SCUDERE, 2015; MAYER - SCHÖNBERGER; CUKIER, 2013; TAURION, 2015).

Demasiadas quantidades de memórias disponíveis, utilizadas e trocadas transformadas em informação, demonstram como o fluxo determina as respostas em cada sistema. Atualmente afirma-se que são cerca de peta bytes de memória para interações diário-mensais de dados acumulados (na proporção de 1 petabyte = 1000 terabytes (Tb), 1 terabyte = 1000 giga (Gb)). Redes Sociais, entre elas o Facebook é a maior responsável por parte, do volume de dados produzidos. (SCUDERE, 2015; TAURION, 2015).

5.2.2 Variedade

A **Variedade** é a dimensão que representa as diferentes tipologias de dados e como podem ser de todos os âmbitos possíveis. São considerados elementos de todas as origens, formatos e de extensões variadas, dados estruturados ou não. É a característica que atribui à diversidade de dados existente dentro dos sistemas, e sua inclusão demonstra como os dados de qualquer tipo de mídia podem ser mesclados e somados a informações produzidas. (IVANOV; KORFIATIS; ZICARI, 2013).

Antes, o que era determinado como informação viria apenas como textos produzidos de origem documental, impressos ou em meio digital. Agora existem mídias de diversos estilos e formatos e extensões que exemplificam e introduzem novas variáveis para análise de dados. Atualmente, tudo o que pode compor a análise no sistema, seja textual ou não, é considerada informação digital. Cada mídia, twitt, postagem, publicação ou e-mail trocado é

agregado e validado pela utilidade com que apresenta nessa era de dados. (IVANOV; KORFIATIS; ZICARI, 2013).

5.2.3 Velocidade

O aspecto da **Velocidade** descreve a capacidade de quão rápido os dados são recuperados, armazenados e processados. Esse atributo é um dos mais significativos, pois devido ao número grande de volume de dados, vindo de vários tipos de dados e de diversas origens, juntamente com uma resposta em uma velocidade singular, faz este atributo ser a junção crucial da tecnologia Big Data. (IVANOV; KORFIATIS; ZICARI, 2013).

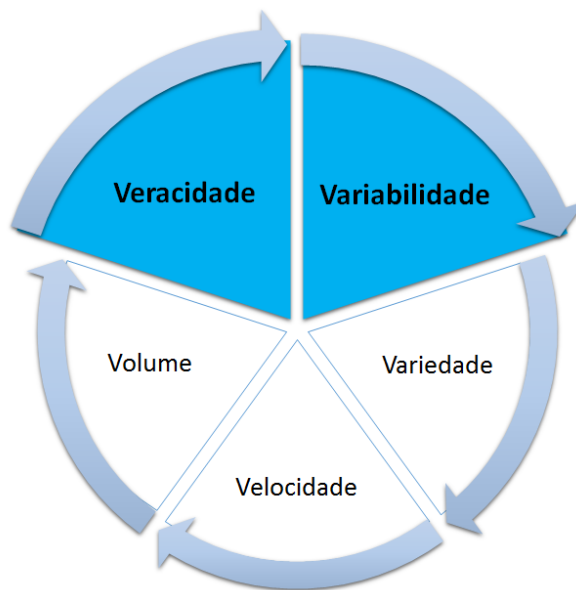
A característica da velocidade é justamente ter a resposta em tempo hábil e de forma precisa, ao compilar o acumulado de informações trabalhando em tempo real e concomitantemente com a variedade delas. Assim como os atributos anteriores, a velocidade do Big Data ser uma tecnologia admirável e essencial atualmente. Existem tecnologias que podem lidar com todas essas variáveis, mas que muitas vezes não é executada corretamente ou nem é possível de processar tão rápido, com a confiabilidade de dados. (IVANOV; KORFIATIS; ZICARI, 2013).

5.2.4 Os 2V's a mais

Há outras duas características fundamentais para consideração de outros parâmetros, mais 2 V's, que vão além das segmentações definidas por Laney (2001). Esses aspectos creditados por outras empresas e autores possuem uma abordagem diferente. Não mais sob uma perspectiva *quantitativa* apresentada nos outros 3V's, mas sim, com um enfoque nas características *qualitativas*. (MAYER - SCHÖNBERGER; CUKIER, 2013; TAURION, 2015).

Essas características qualitativas estão ligadas a como os dados passam a se tornar não só variáveis contadas dentro do sistema. Elas são apuradas para de fato saberem da confiabilidade das informações (**a Veracidade**) e considerar os aspectos que não foram contextualizados dentro da perspectiva naquele momento (**a Variabilidade**). Formam, assim os 5V's, como ilustra a figura 3. (IVANOV; KORFIATIS; ZICARI, 2013; SCUDERE, 2015).

Figura 2 – Os 5 V's



Fonte: (IVANOV; KORFIATIS; ZICARI, 2013)

5.2.5 Variabilidade

A vertente da **Variabilidade** determina as distintas interpretações que os dados podem obter dependendo do parâmetro, ângulo ou ainda quais variáveis são/serão trabalhadas, dentro do contexto específico analisado. Leva-se em conta o que foi considerado dentro dessa primeira análise de apuração de dados. O reflexo que se tem dessa variável transforma a reutilização dos dados viável e útil, uma vez que computadas as informações, elas podem retornar diferentes resultados ou até mesmo uma nova concepção que não vista ou considerada da análise anterior. (IVANOV; KORFIATIS; ZICARI, 2013; TAURION, 2015).

Esse aspecto constrói a ideia de que todo dado por ser utilizado, se não pelo objetivo da pesquisa instantaneamente, mas em algum dado momento. A reutilização do dado em uma pesquisa depende dos critérios que serão determinados de acordo com o ponto de vista pesquisado. Ou o dado usado de forma secundária com uma abordagem diferentemente. Mudando não só o contexto atual da pesquisa, mas aspectos, fatores e parâmetros que possam alterar a visão e análises de cada dado, sem a precisão de uma nova coleta. (IVANOV; KORFIATIS; ZICARI, 2013; SCUDERE, 2015).

5.2.6 Veracidade

A **Veracidade** relaciona-se com a confiabilidade da informação. Pode tanto validar a integridade quanto a precisão dos dados, para assim fazer com que a informação não seja errônea e que, destoa da realidade dos valores analisados. Determinando as diretrizes de cada pesquisa, esse aspecto consegue influenciar um padrão de comportamento analisável dependendo da forma relatada e utilizada e para qual finalidade. (SCUDERE, 2015).

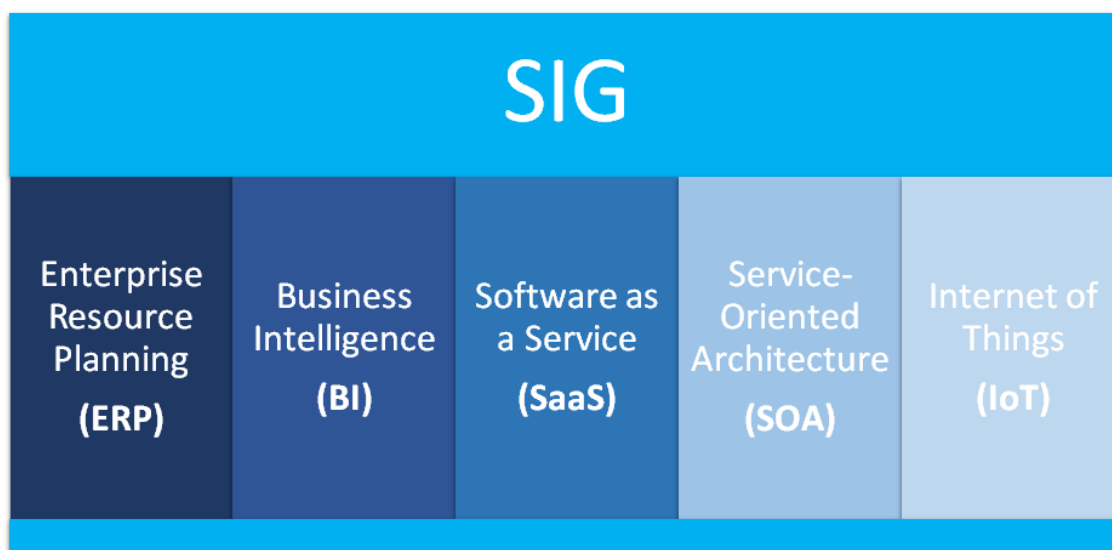
5.3 SISTEMA DE INFORMAÇÃO GERENCIAL

Os sistemas de informação gerencial (SIG) são classificados como o conjunto de recursos de tecnologias voltado para a coleta, armazenamento, processamento e recuperação da informação, transformando os dados em informação dando retorno para os sistemas. (BATISTA, 2004).

Existem aspectos que influenciam pontualmente a estrutura do Big Data na qual a sua tecnologia é fundamentada e/ou relacionada. São *software*, aplicações ou sistemas que definem um conjunto de características aliadas e relacionadas com o processamento de dados. (WINSHUTLLE, 2016). No Quadro 1, foram exemplificadas algumas dessas aplicações. Elas foram escolhidas devido a sua função se assemelhar com a execução do BD, ou o próprio software ser utilizado para embasar o seu funcionamento.

Cada uma dessas tecnologias se interliga motivo pelo qual determinou o porquê de serem destacadas. O *Enterprise Resource Planning*, por exemplo, vem exemplificando a sistematização de vários serviços em uma só plataforma. O *Business Intelligence* é o resultado de toda essa sistematização usada de forma inteligente para a análise e contribuição de tomada decisões na empresa. O *Software as a Service* é a extensão de aplicações que ajudam as empresas a administrar, acessar recursos via *Internet* e o *Service-Oriented Architecture* auxilia a complementar esses serviços, realizando atualizações também pela *Internet*. E por fim, a *Internet of Things* que interligam um dispositivo a outro pela web, sendo facilitada a comunicação de cada um desses softwares.

Quadro 1 - Sistema de Informações Gerenciais



Fonte: (BATISTA, 2004; DAVENPORT, 2014; TAURION, 2015)

5.3.1 ERP (Enterprise Resource Planning)

Uma das tecnologias que foram definidas como base do sistema Big Data começa pela *Enterprise Resource Planning* (ERP). Ele foi desenvolvido como um software que pudesse controlar todos os âmbitos existentes e complexos de grandes organizações empresariais. (RASHID; HOSSAIN; PATRICK, 2002).

O ERP começa em 1980, com uma capacidade reduzida e essencialmente desenvolvida para os aspectos focados para a área de mercado das empresas, aspectos de produção. Como o software estava sendo desenvolvido no auge do aumento dos computadores e, do fenômeno da *Internet*, ele foi redesenhado várias vezes até se adequar as necessidades manifestadas para aquela época. (RASHID; HOSSAIN; PATRICK, 2002).

Em sua definição ele é para gestão de negócios nas empresas, em que envolve todos os processos e subprocessos existentes. Em uma tradução do termo literal, é algo como *Planejamento de Recursos da Empresa*, no qual os aspectos consistem no suporte, planejamento, manufatura de vendas, marketing, distribuição, contabilidade, gestão de recursos humanos, gestão de inventário, gestão de projetos, contabilidade, serviço, manutenção e transporte. (RASHID; HOSSAIN; PATRICK, 2002).

Essa tecnologia permitiu controlar todos os dados que eram manejados na empresa dentro de um só sistema. A revolução do sistema integrado controlando todas as ações da empresa trouxe do ERP a construção de uma das características do Big Data. Mesmo ainda voltado especificamente para gestão de empresa e não para a gestão da informação, a ideia de administração de grandes quantidades de dados, já começava a despertar para os novos meios de processamento de dados. (RASHID; HOSSAIN; PATRICK, 2002).

Até hoje utilizado em vários sistemas de *software*, o ERP possui grande valia diretamente com o Business Intelligence. O modelo gerencial de processos tornou-se fundamental para o BI funcionar. O ERP atua bem no BI devido sua funcionalidade permitir trabalhar conjuntamente com grandes volumes de dados advindos de diversos setores com diversas informações e analisa-los dando base para a resposta e tomada de decisões necessárias. (RASHID; HOSSAIN; PATRICK, 2002).

5.3.2 BI (Business Intelligence)

Não há uma temporalidade fixa para quando o termo apareceu, mas o Business Intelligence (BI) conhecido com o significado de tomada de decisões foi definido por Gartner Group em 1990. Contudo, desde 1970 o mundo dos computadores estava se direcionando para o que de fato a tecnologia abrange. (LIMA; LIMA, 2011; SCHEPS, 2008).

Programadores começaram a desenvolver sistemas de gerenciamento em banco de dados (SGBDS), procurando lidar com a grande estrutura de dados que se ampliou com a criação dos computadores. Com as grandes demandas de dados crescendo, os sistemas de análises foram se tornando melhores e poderosos. Assim, a solução adotada foi unir o sistema, para gerenciar as operações diárias das empresas. A versatilidade das transações diárias sendo analisadas conjuntamente, como tarefas simples de representantes de vendas, ao cobrar o pagamento e emitir o recibo ao cliente, transformou-se em um módulo conectado em reunir e organizar informações das vendas para que posteriormente pudessem ser analisado-utilizadas. E com isso trazer o modelo inteligente BI. (SCHEPS, 2008).

O BI é considerado uma gama de soluções tomadas a partir de uma análise previamente considerada e feita de acordo com os desdobramentos de uma empresa. É a apuração e análise de quantidades vastas de dados a fim de se obter respostas estratégicas, para a tomada de decisões, tendências e previsão comerciais de acordo com as informações adquiridas. Sua proposta é analisar os dados e depois transformá-los em informações que deem embasamento para a conversão em conhecimento, úteis ao sistema. (HAHN, 2000; GARTNER, 2016).

Voltado para os negócios, o BI fundamentou-se em processos e subprocessos responsáveis pelas análises de dados de cada setor ou esfera dentro de uma empresa. Com pensamento nisso, foram qualificados diferentes aspectos de análise e processamento, de informação no sistema a fim de se obter soluções, devido a cada tomada de decisões. Dentre esses aspectos, existem componentes que contribuem para a infraestrutura, são o Data Warehouse e Data Mining. (REINSCHMIDT; FRANCOISE, 2000).

5.3.2.1 Data Warehouse (DW)

Como um dos componentes do funcionamento do Business Intelligence, o Data Warehouse tem a função de garantir que os bancos de dados continuem online a fim de fazer com que as informações estejam sempre disponíveis para serem analisadas e processadas. A ideia principal de um DW é fornecer uma infraestrutura que contenha todas as informações dos sistemas operacionais da empresa, incluindo dados históricos acessíveis. Transformando-se em relatórios e consultas. (REINSCHMIDT; FRANCOISE, 2000).

5.3.2.2 Data Mining

O Data Mining é responsável por ser um processo de descoberta de informações a fim de transformá-las em dados úteis e reaproveitáveis no sistema. A informação extraída pode ser retirada a fim de identificar um método que acompanhe as relações de registro do banco de dados ou crie uma previsão. Dependendo do método utilizado para extrair essas informações, criam-se diferentes informações obtidas através dos registros. (REINSCHMIDT; FRANCOISE, 2000).

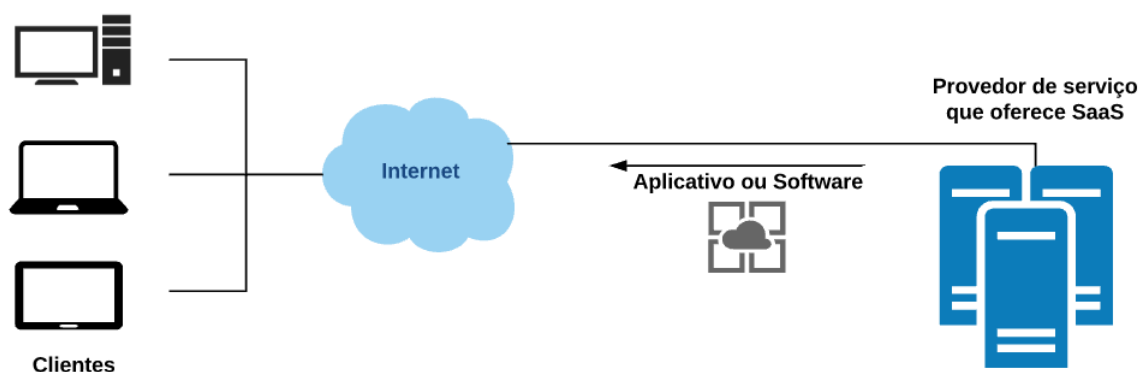
O benefício do BI é trazer esse sistema integrado que facilita a administração dos dados, de grandes quantidades e diferentes esferas para uma estruturação pautada na organização das informações. Tornando as decisões respaldadas pela as respostas que o sistema pode trazer, por meio das análises feitas previamente dos dados obtidos. Fazendo assim, economizar custos, diminuir os riscos e obter decisões táticas em curto período possível, na melhor hipótese, em tempo real. (REINSCHMIDT; FRANCOISE, 2000; LIMA; LIMA, 2011).

5.3.3 SaaS (Software as a Service)

Com a *Internet* como canal de comunicação e informações surgem diversas tecnologias que começam a prestar inovadoras atividades. Um dos mais tradicionais e conhecido como tecnologia de computação em nuvem (Cloud Computing) é o *Software as a Service* (SaaS), um software como serviço. Uma tecnologia desenvolvida desde a década de 2000 que tem sido explorada atualmente, e com sua diversidade de aplicações tornou-se uma poderosa ferramenta. (CHEE; FRANKLIN JUNIOR, 2013).

O *Software as a Service* (SaaS) é um aplicativo que é hospedado em um servidor remoto e acessado pela a *Internet* como um serviço, como pode ser visualizado na Figura 4. O provedor do aplicativo é o que o mantém e é o responsável pelas alterações e aprimoramento necessários do software. Um exemplo é o e-mail ao ser acessado através da *Internet*; a Microsoft oferece o Hotmail e fornece acesso ao programa Microsoft Word (editor de texto da Windows) de forma online. Tanto o e-mail quanto o programa são considerados SaaS. (VELTE; VELTE; ELSENPETER, 2013).

Figura 3 – Fluxograma SaaS



Fonte: (VELTE; VELTE; ELSENPETER, 2013)

O que o difere é por ser um serviço especificamente para uso de ferramentas da Web. E o que o distingue do Cloud Computing é que o SaaS não fornece em si realmente o armazenamento, ele é um canal intermediário que terceiriza o serviço. Empresas que fornecem o Cloud Computing oferecem seus provedores para serem “alugados”. O que é vantajoso no SaaS, ao pagar pelo software sai mais em conta pelo fato de alugar momentaneamente o serviço, do que uma empresa que oferece diretamente o armazenamento em cloud computing, no qual o custo é alto. (VELTE; VELTE; ELSENPETER, 2013).

Existem desafios a serem ultrapassados pelo SaaS, a sua arquitetura ainda limita o uso para grandes empresas que necessitam de um grande fluxo de informações para seus clientes. As aplicações não se adequam ao estilo do BI, por exemplo. As transições de informações trocadas entre clientes e servidor exigem do cliente uma banda larga e CPU (unidade central de processamento) que suporte uma infraestrutura para o processamento das solicitações existentes. (VELTE; VELTE; ELSENPETER, 2013).

5.3.4 SOA (Service-Oriented Architecture)

Outro serviço ligado ao cloud computing é o Service-Oriented Architecture (Arquitetura orientada a Serviço), que é responsável em atualizar as tecnologias dos *software* que oferecem os serviços. Ele unifica e otimiza os processos de negócios por meio de composição de grandes aplicações como um conjunto de serviços. (VELTE; VELTE; ELSENPETER, 2013)

O benefício de suas funções é oferecer uma plataforma de computação orientada para serviços, introduzindo novas tecnologias e determinado a evolução das soluções ofertadas. Fazendo assim, se ter mais opções de fornecedores e mais retorno para os investidores. (VERAS, 2015).

Assim como o SaaS o SOA vem sendo desenvolvido em prol da operabilidade e manutenção de *software* de *Internet*. O que há em comum entre eles é que usam um modelo de serviços desenvolvidos para conexão entre o data center e servidores. E o que os difere é que a SOA é responsável pelo *design* e automação dos processos de negócios. Enquanto o SaaS busca oferecer o *software* do serviço em si. (VELTE; VELTE; ELSENPETER, 2013).

5.3.5 Internet of Things (IoT)

A *Internet of Things* (IoT) ou *Internet* das Coisas surgiu seu conceito em uma palestra em 1999 cunhado por Kevin Ashton (2009). Na época, Ashton nomeava tudo que pudesse acessar a rede por meio de qualquer tipo de dispositivo, era a *Internet* de tudo. Começou-se a especular que o futuro da tecnologia seria dominado por um conjunto de “coisas” que seriam controlados pela *Internet*. Uma rede onde faria com que os dispositivos e máquinas pudessem interagir ou ser controlada entre si a longa distância, apenas por meio da *Internet*. (ASHTON, 2009; LEE; LEE, 2015).

Dez anos depois Ashton (2009) defendia ainda que a IoT pode e vai mudar a forma como os humanos constroem seu mundo. Afirma que, por mais que as necessidades humanas sejam diferentes do que dez anos atrás, a valorização das “coisas” tem alcance real e potencial assim como a *Internet*. Mudará não só a forma com que as pessoas convivem com a tecnologia, mas também alcançará todos os âmbitos sociais. Ainda que em transição, a tecnologia das “coisas” já determina as possíveis perspectivas do futuro. (ASHTON, 2009).

A IoT é conceituada como dispositivos que possam ser conectados entre eles através da *Internet*. A interação existente entre **dispositivos → Internet → humanos** revela como funciona a tecnologia da *Internet* das coisas. Mostra como a capacidade e o impacto informacional existente, de qualquer tipo de aparelho que, ligado à *Internet* e com configurações relacionais entre sistemas de computadores consegue conectar a outros tipos de aparelhos. As

aplicações aparecem dos mais diversificados estilos, e com elas são possíveis listar como e quais parâmetros são associados aos diferentes aspectos de tecnologia. (LEE; LEE, 2015).

A IoT se baseia em 5 produtos e serviços que interligam os dispositivos. São eles, *radio frequency identification* (RFID), *wireless sensor networks* (WSN), *middleware*, *cloud computing* e *IoT application software*. Essas ferramentas serão eficientes quanto uma análise de vários dados significativos em tempo real, e em como as pessoas estão tomando decisões. Compartilhado e colaborando para a fomentação da informação de forma ampla para o sistema, otimizando os serviços e produtos prestados. (LEE; LEE, 2015).

Radio Frequency Identification (RFID) é um método que permite identificação automática por meio de radiofrequência na qual a etiqueta (*tag*) utilizada possui dados armazenados e podem ser acessados por meio de um sensor chamado leitor (*reader*). Usualmente é uma tecnologia que suporta mais dados que os tradicionais códigos de barras. Os microchips utilizados podem ser acessados, modificados e recuperados através do leitor (*reader*). (LEE; LEE, 2015).

Wireless Sensor Networks (WSN) é dispositivo com sensores distribuídos capazes de monitorar aspectos ambientais e condições físicas do estado das coisas e que podem cooperar com os sistemas de RFID. Acompanhando os estados das coisas, como localização, temperatura, situação, movimentos etc. (LEE; LEE, 2015).

Middleware é uma camada de software responsável por controlar a comunicação entre softwares e demais aplicações. Tornando mais fácil o transporte de informações e dados entre diferentes plataformas e sistemas. Facilitando também a conversação entre os mais diversos dispositivos, e o retorno sobre a qualidade dos programas desenvolvidos da IoT. (LEE; LEE, 2015).

A computação em nuvem é um modelo no qual é evidenciado como funciona a IoT. Ligada diretamente na *Internet*, ela presta serviço diretamente através de serviços disponíveis que facilitam a comunicação entre rede e software. Essa combinação gera uma quantidade razoável de dados, e manipulação de dados conectados facilitando um fluxo de informações entre *Internet*, aplicação e usuário. (LEE; LEE, 2015).

A *IoT Application Software* facilita o desenvolvimento de uma indústria voltada para usuário e máquinas. Suas aplicações habilitam o dispositivo a outro dispositivo e uma interação de humanos com outros dispositivos, criando uma conectividade física de forma confiável e confidencial. Em um tempo hábil de forma adequada para cada necessidade apresentada. (LEE; LEE, 2015).

5.4 TECNOLOGIAS UTILIZADAS

A arquitetura da informação que foi desenhada através da nova "*datificação*" (transformar tudo em dados) demonstra como as novas tecnologias conseguiram se adequar ao funcionamento do BD, ou como foram desenvolvidas com ele. A inserção das máquinas no contexto de Big Data mudou tanto que, a possibilidade de vários modelos lidarem com a proposta de grandes quantidades de dados para análise, se encaixa de forma que os dados são otimizados e previstos. (DAVENPORT, 2014).

Essas tecnologias são as principais, devido ao processamento de dado ser algo incutido nelas. São da geração de processamento de dados com a questão do uso das máquinas, onde foi implementado e trazendo o grande número de fluxo de dados. Tornando assim, uma nova geração de informações construídas de acordo com a necessidade de informação. Cada uma dessas tecnologias atende a aspectos diferentes do BD e cada uma é responsável por trabalhar com o quantitativo de dados existentes. No Quadro 2, são especificado algumas dessa tecnologias. (DAVENPORT, 2014).

Quadro 2 - Visão geral das tecnologias de Big Data

Visão geral das tecnologias de big data

Tecnologia	Definição
Hadoop	Software de código aberto para o processamento de big data em uma série de servidores paralelos
MapReduce	Um framework arquitetônico no qual o Hadoop se baseia
Linguagens de script	Linguagens de programação adequadas ao big data (por exemplo, Python, Pig, Hive)
Aprendizado de máquina	Software para identificar rapidamente o modelo mais adequado ao conjunto de dados
Visual analytics	Apresentação dos resultados analíticos em formatos visuais ou gráficos
Processamento de linguagem natural (PLN)	Software para análise de texto – frequências, sentido etc.
In-memory analytics	Processamento de big data na memória do computador para obter maior velocidade

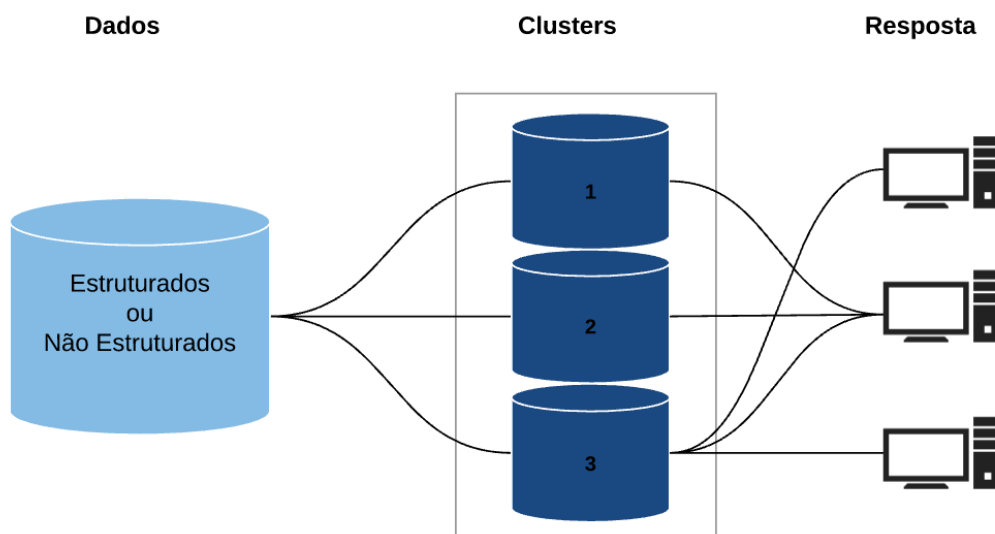
Fonte: (DAVENPORT, 2014)

5.4.1 Hadoop

O Hadoop foi desenvolvido pela Apache desde 2005 e foi lançado em 2011. Inspirado no que a Google fez com o MapReduce, o Hadoop tem sua estrutura baseada no mesmo estilo de processamento de dados. (HADOOP, 2016; TAURION, 2015).

O Hadoop é um projeto cujo software livre, processa grandes conjuntos de dados estruturados ou não, a partir de um único servidor, distribuídos em clusters (aglomerados) para armazenamento, com consequência de respostas dependendo da análise feita. Como pode ser visualizado pela Figura 5. (HADOOP, 2016).

Figura 4- Como funciona o Hadoop



Fonte: (HADOOP, 2016)

Ele ainda se divide em 4 módulos: *Hadoop Common*, que se refere a bibliotecas e utilitários comuns que suportam os módulos do Hadoop; *Hadoop Distributed File System* (HDFS), um sistema de arquivos distribuídos que fornece acesso de grande produtividade para os dados do aplicativo; *Hadoop YARN*, uma estrutura que escalona as tarefas e gerencia os recursos em clusters e *Hadoop MapReduce*, que se utiliza da YARN, para processar os grandes conjuntos de dados. (HADOOP, 2016).

O Hadoop é uma das mais importantes tecnologias desenvolvidas do BD. Por ser um projeto *Open Source* (Código aberto para livres alterações e comercialização) permite criação de sistemas específicos para a necessidade de cada empresa. Com a popularização de cuidar de grandes dados não estruturados, o Hadoop é o que mais se adequa e o mais mutável dos *software*. (TAURION, 2015).

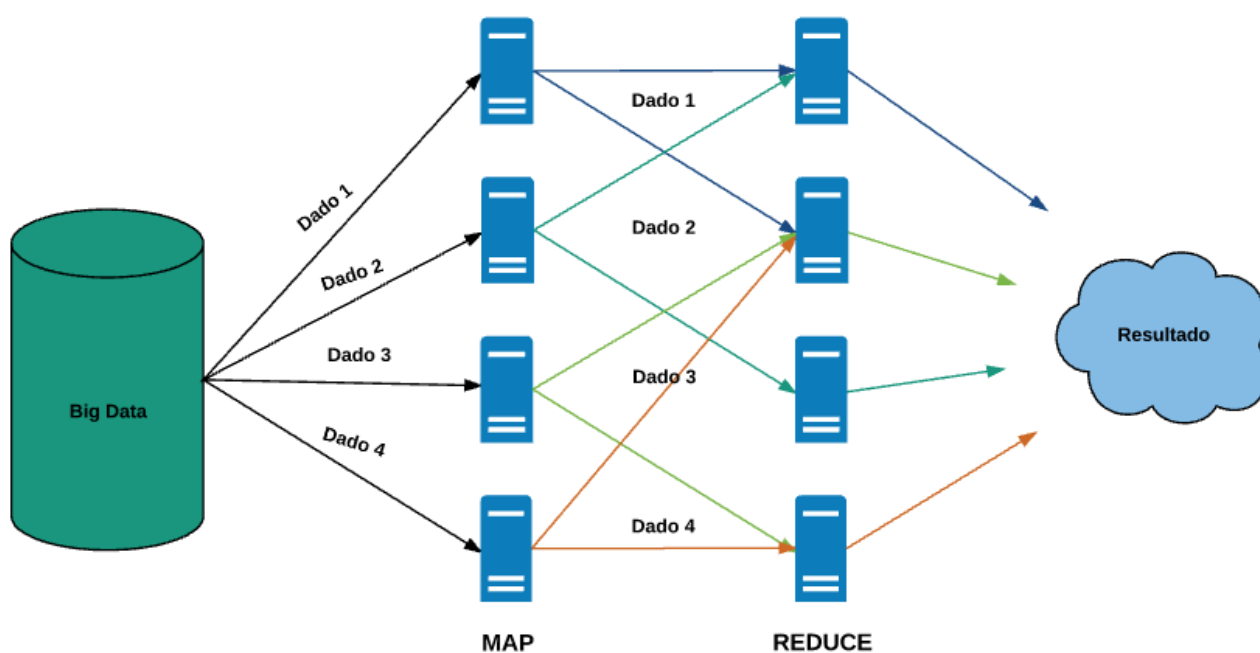
5.4.2 MapReduce

Desenvolvido pela Google em 2004, o MapReduce foi apresentado por um artigo publicado de autoria de Jeffrey Dean e Sanjay Ghemawat. No artigo, eles apresentaram a ideia de como os dados poderiam se tornar aglomerados e como essa arquitetura traria a facilidade de processamento nessa estrutura. (DEAN; GHEMAWAT, 2004).

O MapReduce é um modelo de programação para processamento de grandes conjuntos de dados. O programa se divide em duas funções, a função de *map* (mapa) e a função de *reduce* (redução). Essas funções recebem um valor chave/par para gerar outros valores chave/par que mesclam com outros valores chave/par. O que acontece é que a opção *map*, mapeia por algo e a opção *reduce* mescla e resume os resultados conforme os valores intermediários aparecem. Como pode ser visto na Figura 6. (DEAN; GHEMAWAT, 2004).

O modelo deu tão certo, que o MapReduce serviu de base da tecnologia do Hadoop. Ele conseguiu definir como funcionaria o sistema de MapReduce e consequentemente ocasionou para as funcionalidades de outros *software*. (DAVENPORT, 2014).

Figura 5- Como funciona MapReduce



Fonte: (DEAN; GHEMAWAT, 2004)

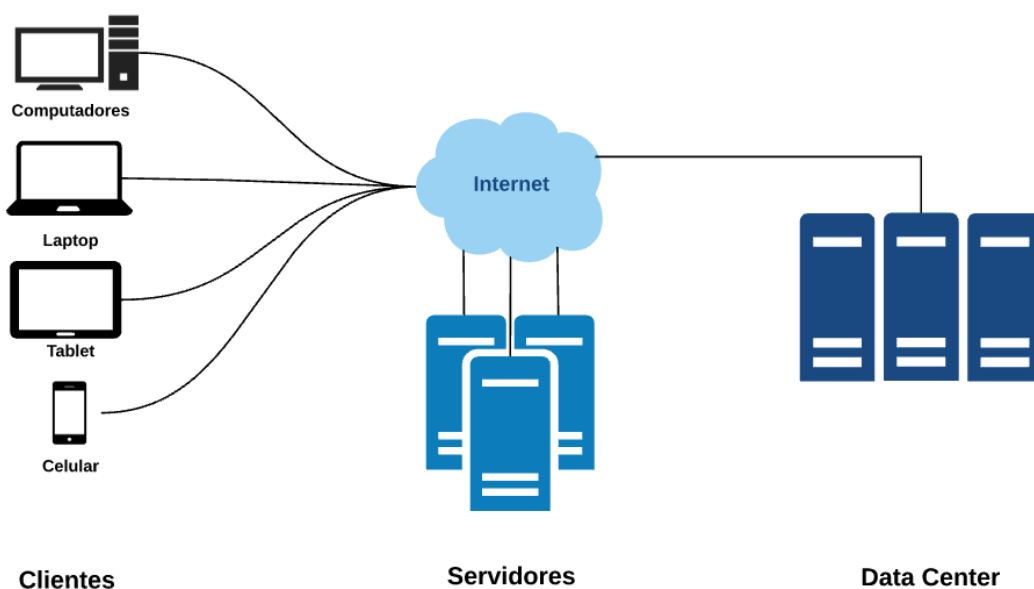
5.4.3 Cloud Computing

Com o uso mais comum da *Internet* nas atividades diárias das pessoas, começou-se a pensar em uma forma de deixar serviços e ferramentas hospedadas diretamente na *Internet* de forma a facilitar o acesso. Foi com essa ideia que nos anos 2000, começaram a surgir os serviços de armazenamento em nuvem. E com o decorrer dos anos, tornou-se como um serviço mais

comum e acessível a todas as pessoas e empresas que necessitavam de atividades de forma rápida por intermédio da *Internet*. (CHEE; FRANKLIN JUNIOR, 2013).

A Computação em nuvem é definida com qualquer serviço ou atividade que, por intermédio da *Internet* consiga criar conexão ou transmissão de dados entre data centers, servidores e usuários. Partindo dessa premissa, determinam-se diferenciadas funções que possam prestar todo e qualquer tipo de serviço pela *Internet*. Como pode ser demonstrada pela Figura 7. (CHEE; FRANKLIN JUNIOR, 2013).

Figura 6- Cloud Computing



Fonte: (VELTE; VELTE; ELSENPETER, 2013)

O cloud computing foi durante muito tempo definido como uma tecnologia abrangente de tudo relacionado à *Internet*. Mas sua definição com os anos foi tornando-se mais delimitada e sua conceituação tornou-se mais consolidada. E com um entendimento mais conciso, Cloud Computing é determinado como a capacidade de memória e armazenamento entre computadores e servidores pela *Internet*. (VELTE; VELTE; ELSENPETER, 2013).

O cloud computing é responsável por uma escala de diversos serviços oferecidos com aplicações via *Internet*. Dentre esses serviços, pode ser ofertado qualquer tipo deles e acessados de qualquer local, independentemente da plataforma utilizada. Algo prático, sem precisar ser instalado, mas utilizando como se fosse. (VELTE; VELTE; ELSENPETER, 2013).

A arquitetura do cloud computing depende de alguns elementos responsáveis e atuantes para o seu funcionamento, esses elementos são: clientes, servidores e data center. Os clientes geralmente são computadores, celulares e laptops, são dispositivos que poderão se conectar a nuvem. Os servidores que, são a estrutura mantenedora da *Internet*. E por fim, o data center, que é onde fica armazenado aplicativo, que são o conjunto de vários servidores também. (VELTE; VELTE; ELSENPETER, 2013).

Como citado em itens anteriores, uma das aplicações do Cloud Computing mais usuais é o SaaS e SOA. São aplicações que interagem com o tipo de serviço entre data center e servidores/computadores. Basta acessar a *Internet* que é possível controlar todos os documentos e/ou dados armazenados na nuvem. (VELTE; VELTE; ELSENPETER, 2013).

Os benefícios de se manter o gerenciamento de dados através do cloud computing é principalmente pelo fator armazenamento. É possível fazer acessos de todos os dados, de qualquer lugar e independentemente da plataforma, basta se ter acesso à *Internet*. E é assim que vem se relacionando com o BD, pois de forma analítica consegue ter um retorno imediato e de certa forma, manipular os dados. (VELTE; VELTE; ELSENPETER, 2013).

O serviço oferecido depende de uma terceira entidade onde sustenta a tecnologia. A terceira entidade é quem determina a quantidade de memória e a velocidade. Traz a preocupação quanto pela acessibilidade dos dados, levando-se em conta que todos os dados ficaram armazenados, mas que talvez nem sempre estejam disponíveis. (VELTE; VELTE; ELSENPETER, 2013).

O serviço pode ainda ser facilmente interrompido. Se houver problemas com quem oferta o armazenamento ou os servidores de ambas as partes não comunicarem entre si pela eventualidade de falta de *Internet* ou de luz, por exemplo.

5.4.4 Analytics

O analytics tradicional começa se confundindo com o BD, devido seus conceitos estar intimamente vinculados. Não havia tanto alarde quanto a questão do Big Data, pois as empresas e cientistas já achavam estar lidando com ele. Desde os anos de 1950 pelo menos, o analytics já tinha ferramentas sendo utilizadas no universo dos negócios e desde então não tem sido diferente, apenas sendo moldadas as novas realidades. (DAVENPORT, 2014).

Cada aplicação do analytics está relacionada com as novas tendências e tecnologias que apareceram ao longo do tempo. Existem eras que designa os atributos para cada época. Essas eras são nomeadas como Analytics 1.0, Analytics 2.0 e Analytics 3.0, exemplificadas pelo o Quadro 3. (DAVENPORT, 2014).

Quadro 3 – Analytics

	Analytics 1.0	Analytics 2.0	Analytics 3.0
Tipos de empresas	Grandes empresas	Startups e empresas on-line	Todas –“economia dos dados”
Objetivo do analytics	Decisões internas	Novos produtos	Decisões e produtos
Tipo de dados	Pequenos, estruturados	Grandes, não estruturados	Todos os tipos combinados
Abordagem de criação	Ciclo longo, em batches	Ciclo curto, ágil	Ciclo curto, ágil
Tecnologia básica	Pacotes de software	Software livre	Amplo portfólio
Principal tipo de analytics	Descritivo	Descritivo, preditivo	Prescritivo
Relacionamento com o lado do negócio	Função de apoio	“Na ponte de comando”	Colaborativo

Fonte: (DAVENPORT, 2014)

5.4.4.1 Analytics 1.0

A era Analytics 1.0 começou de 1954 e ficou até 2005. As características que definem essa era condizem pelo fato de as fontes de dados, serem pequenas e internas. Ou seja, muito pouco dado e **coleta** dele de dentro da empresa/sistema. Outra característica é o fato dos dados serem armazenados antes mesmo de serem analisados. E que sua atividade analítica, consistia

apenas na produção de relatórios. Suas análises quantitativas eram separadas das decisões e dos analistas com a função de apoio, e às vezes, as empresas nem chegava a utilizar esses dados em seus benefícios. (DAVENPORT, 2014).

5.4.4.2 Analytics 2.0

Considerando o ano 2000, como o início do Analytics 2.0 juntamente com o nascimento do Big Data, essa era já está voltada para as novas explorações de dados online, e empresas também. Não havia só como atenção, como na Analytics 1.0, as decisões internas, mas também os serviços e produtos oferecidos aos clientes. Mas suas atividades analíticas em compensação continuavam voltadas para os dados internos. Onde tinham dados sobre o cliente, mas não utilizava como comparação de análises com outros dados. (DAVENPORT, 2014).

Ainda na era Analytics 2.0, os dados obtidos dos clientes começavam a dar indício à falta de estrutura com a quantidade de fluxo que ocorria. Dados não estruturados e com muito volume, mostram que ainda afetava as atividades analíticas. Com velocidade da análise sendo rápidos, os cientistas começam a querer estar à frente, devido às respostas serem rápidas e úteis em algum ponto do sistema. (DAVENPORT, 2014).

5.4.4.3 Analytics 3.0

Já a era Analytics 3.0, mostra que atualmente o mundo se encontra em processo de transição, entre a Analytics 2.0 e a 3.0. Afirma-se que a Analytics 3.0 será uma junção do melhor das eras 1.0 e 2.0. Mas que ainda sim, possui uma característica qualitativamente diferente das outras eras. Seria algo como uma mescla entre Big Data e Analytics tradicional com velocidade e impacto, gerando grandes ofertas e insights para qualquer empresa e seus serviços. (DAVENPORT, 2014).

5.5 APLICAÇÕES DO BIG DATA

A tecnologia big data tem influenciado todos os setores e organizações. Ao entender como o Big Data funciona, compreende-se a extensão de como ele se adequa a uma sociedade voltada para uma renovada arquitetura de informação. O BD tem transformado as formas naturais de trabalho atuais e

com isso uma adaptação aos novos instrumentos que são necessários para ocorrer tal mudança. Como será a graduação dessa mudança e como as organizações se manifestarão, será crucial para a resposta ao sistema e para a própria organização. (DAVENPORT, 2014).

Os sistemas são, criados para se aprimorarem com o tempo analisando por meio de sinais e padrões encontrados, soluções que transformem os dados em algo útil e para uso real e concreto. Assim, eles seriam sempre baseados em medidas aplicáveis, desenhada pelo seu próprio sistema. (MAYER - SCHÖNBERGER; CUKIER, 2013).

E mesmo que o sistema seja aperfeiçoado, o desafio real aponta para como será elaborado os padrões de dados e como serão reunidos a fim de, deixá-los úteis. É nesse estímulo que ele afirma que, os cenários previstos dos setores serão baseados de acordo como cada organização lida com os dados e se os possui. (DAVENPORT, 2014).

Ao averiguar a maneira que as organizações operam seus dados, classificas em 3 estilos. Em empresas líderes, onde seu desempenho excedem as expectativas. Um exemplo é a Netflix e a Amazon, são companhias que baseadas nas preferências de seus usuários, serviram para o processamento dados, convertidos em vários padrões de comportamento e se consagraram na previsão de oferta de serviços. (DAVENPORT, 2014).

As empresas desprovidas de dados, as quais não os tinham ou não eram bem estruturados, como por exemplo, as organizações de saúde onde mesmo que as anotações dos históricos dos pacientes estivessem em prontuários on-line, o texto não estruturado tinha déficit de anotações relativo aos pacientes, atrapalhando o desenvolvimento da instituição. (DAVENPORT, 2014).

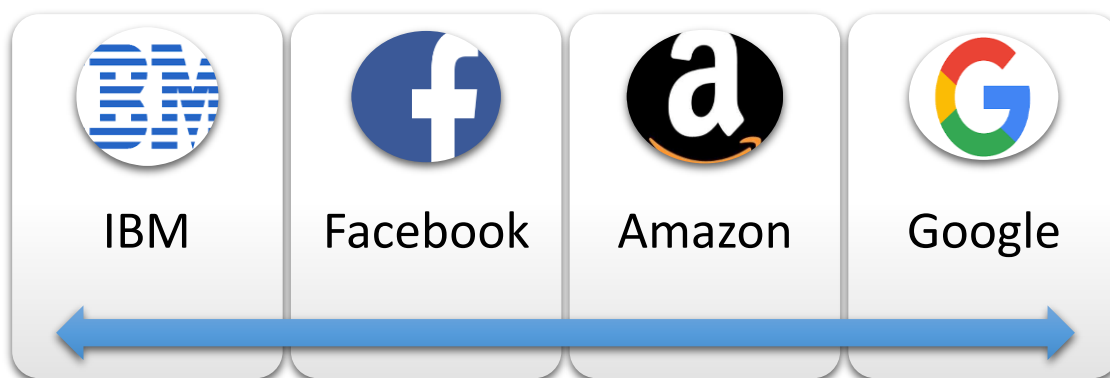
E por fim, as empresas que não utilizavam os dados para seu próprio benefício e nem dos seus clientes. Instituições como bancos, que utilizavam os dados sobre a movimentação bancária de cada cliente apenas para interpretar e oferecer serviços de marketing. (DAVENPORT, 2014).

A primeira atividade realizada em big data foi protagonizada por empresas de produtos e serviços de *Internet*, e em startups que atuam na *Internet* e em setores semelhantes. Parte dessas empresas que contribuem são essenciais para a consolidação da ciência de dados. Cada avanço tecnológico foi devido a

essas várias empresas existentes. Há algumas delas que são de notório sucesso é que afetam diretamente a natureza do BD. (DAVENPORT, 2014).

São entidades significativamente catalisadoras da nova ordem de dados. Instituições como IBM, Facebook, Google e Amazon vista na Figura 8 são modelos tanto de como o que era o antigo transitou para a o atual; e como empresas que surgiram desses novos cenários conseguiram se estabelecer. Partindo dos exemplos que acompanham o surgimento da criação de cada empresa, um estudo levantado consegue estender a importância de cada entidade devido a sua estrutura única, ser parâmetro de tecnologias diferenciadas que acrescentam cada uma em seu estilo, contribuições de tecnologias atuais.

Figura 7 - Aplicações do Big Data nas empresas



Fonte: (DAVENPORT, 2014)

5.5.1 IBM

Uma empresa focada na tecnologia, a IBM desde 1880, vem se estruturando para a nova ordem mundial de conhecimento. Com os avanços da tecnologia, ela foi determinando seu espaço e contribuindo para a formação de nova informação que foi despertando no mundo. (IBM, 2016).

Com a Tabulating Machine Company de Hollerith, em 1896, a IBM alavancou como a empresa de desenvolvimento de base tecnológica que movimentou as formações dos computadores e informações e que, até atualmente é referência de tecnologia no mundo. (IBM, 2016)

A IBM é uma organização consagrada por transformar seus segmentos de mercado e investir em seus profissionais trazendo sempre ideias inovadoras. Dentre esses investimentos, no meio de tantas ofertas estimuladas por uma nova corrente de tecnologias, a IBM se envolveu rapidamente com a ideia de processamento automático. E é por meio da própria, que se obtém o preâmbulo de uma nova era de dados. (IBM, 2016)

Por ser uma empresa que se intitula como “empresa de tecnologia da informação do mundo”, a IBM se adequa aos novos embates de globalização, de forma que suas tecnologias sejam agentes de novas propostas. Acontece que, com a liberação de software atual, como a Apache fez com o Hadoop, permitiu a IBM se alinhar e escrever seus próprios estilos de manuseamento de dados. Criando banco de dados relacionais e multiplataformas em servidores assegurados por cloud computing. A IBM é uma das principais referências em software para o controle de informações de uma empresa. (DAVENPORT, 2014).

A IBM defende que com o Big Data pode-se fazer o que quiser da forma que quiser. As diferentes quantidades de dados coletados podem gerar visões e resultados incríveis porque enriquece as iniciativas de análise que estão ocorrendo nas empresas atualmente. (ZIKOPOULOS et al., 2015).

Com a plataforma de infraestrutura oferecida pelas bases de dados relacionais, um método similar com as formas de gerenciamento de dados já existentes. A IBM consegue transcender os meios de processamento e cria um estilo baseado nas características dos softwares existentes. Integrando-se os métodos do BI e o ERP, a IBM utiliza-se dessas ferramentas para fundamentar sua forma de trabalhar e oferecer amplos serviços de gerenciamento. (ZIKOPOULOS et al., 2015).

A IBM explica que a melhor maneira de utilizar o Big Data e seus serviços de gerenciamento é entender a origem dos dados, como fazer para analisá-los e aproveitá-los de forma que gere resultados analíticos e que consequentemente resultem insights para dentro da empresa. É dessa forma, que a empresa afirma, vem se adequando as demandas informacionais reais e atuais. (ZIKOPOULOS et al., 2015).

5.5.2 Facebook

Criada em 2004 por Mark Elliot Zuckerberg e cofundadores o Facebook começou quando Mark ainda frequentava a Universidade de Havard. A ideia era criar uma conexão virtual entre as pessoas que conviviam pelo campus da universidade. Assim, as pessoas que ingressassem no “TheFacebook” (chamado antigamente, na época do lançamento) criariam um perfil online onde seria possível colocar todas as informações pessoais e profissionais como desejassem. O site vingou tanto que apenas em alguns meses havia se disseminado e várias outras universidades aderiram à ideia. (FACEBOOK, 2016).

O Facebook é uma empresa que gerencia milhões de dados e é elevada a quantidade de mídias manipuladas diariamente. O intuito do Facebook é fazer com que as pessoas se conectassem mediante uma rede, onde interligaria tudo em tempo real. Voltada para a comunidade ativa na internet, ele explora diversos meios de serviços e abusa dos produtos oferecidos. (FACEBOOK, 2016).

Tais produtos e serviços oferecidos são informações online disponíveis e acessadas a qualquer hora e em qualquer lugar. Atualizado e feito para ser exibido em qualquer plataforma e carregar qualquer tipo de mídia. O site suporta diferentes formatos como fotos, vídeos, links, gifs e textos entre outras formas e formatos caracterizados pela a rede social. O que até então em nenhuma outra mídia conseguia oferecer. Essas múltiplas opções em uma mídia só acrescenta como a melhor rede social. (FACEBOOK, 2016).

O que consagrou o Facebook foi o fato de ir a favor da tecnologia juntamente com a interatividade pessoal. Ao apostar nos interesses pessoais e sociais e estar sempre de olho no que os usuários publicam, escrevem e compartilham, fez da rede social ter uma análise contínua de todas as tendências de comportamento possíveis. Conseguindo definir um padrão de pensamento, vontades e necessidades e saber da predileção de cada usuário no Facebook conseguiu desencadear um novo tipo de estreitamento de serviço tanto entre as marcas quanto entre as empresas e pessoas integradas. (DAVENPORT, 2014).

Por meio dos dados coletados, as informações contribuem para avaliação recorrente de algum serviço oferecido de uma determinada empresa

ou a qualidade de um produto. Ainda muito utilizado para criar essa visibilidade e acessibilidade entre esses diferentes usuários, o Facebook consegue fazer seu negócio perdurar. (DAVENPORT, 2014).

Um fator relevante é quanto a sua confidencialidade. É claro que o usuário controla o que pode ou não ser visto dentro da sua homepage individual. Mas a quantidade de dados que são gerados e codificados de forma analiticamente exposta pelo Facebook sucintamente, não traz a quantidade de exposição de suas informações. (DAVENPORT, 2014).

É claro que a empresa traz questões de privacidade preservadas e faz questão de notificar o usuário quanto à pretensão do uso de dados. Mas a verdade é que não se sabe realmente como esses dados são/serão utilizados. Mesmo ao concordar com o fornecimento de dados. Não se sabe o grau de privacidade real está que sendo controlado. E nem se são usados com um propósito “inofensivo”. (DAVENPORT, 2014).

5.5.3 Amazon

Uma organização responsável por vender produtos via internet, como livros, aparatos tecnológicos, utensílios gerais entre outras coisas. A Amazon é considerada a revolução do mercado e marketing digital. Seu alcance a nível mundial a consagra como a melhor loja virtual de vendas, direcionada aos clientes online. (DAVENPORT, 2014).

Acompanhado a revolução da *Internet*, Jeffrey Bezos teve a ideia de criar um site de vendas enquanto trabalhava em uma grande empresa. Imaginou um novo negócio em que se baseava em produtos oferecidos online. Então 1994, resolveu fundar a Amazon, criando o conceito de mercado online para certos tipos de artigos. (FUNDABLE, 2017).

Por ser uma organização criada dentro dos parâmetros de uma diferente modelagem de dados, ela consegue estruturar um serviço onde as tendências possam ser controladas e observadas pelas preferências coletadas de dados da própria navegação do usuário. Quando se observa, por exemplo, o que o usuário pesquisou e se comprou, a partir desse ponto, são analisadas outras variáveis de linhas sugestivas de uma nova possível compra, ou de recomendação que poderá ser sugerida no futuro. Isso demonstra como a

quantidade de dados coletados foi transformada, nesse sistema, uma nova proposta de marketing. (DAVENPORT, 2014).

A Amazon se consolida dando manutenção em seus negócios constantemente, devido à sua arquitetura única, baseado em cloud computing criando seu banco de dados sobre os produtos. Constrói seu domínio na apropriação desses recursos, de forma singular. A premissa dessas tecnologias, como elas são organizadas e elaboradas, casam com a medida de sucesso que a fundamentou. (DAVENPORT, 2014; VELTE; VELTE; ELSENPETER, 2013).

O sistema de recomendação (também utilizado nas ferramentas da Google) demonstra como as análises de dados, no caso a pesquisa de produtos, funcionam muito bem a partir de variáveis inseridas pela predileção de uma pessoa. Mostrando assim como o princípio do Streaming Service, que até então era pouco comum, se torna ferramenta chave para a projeção de excelentes serviços de compra e de divulgação. (DAVENPORT, 2014; VELTE; VELTE; ELSENPETER, 2013).

A abrangência de sua tecnologia engloba tanto a infraestrutura quanto o produto final. Por ser o serviço mais amplo de nuvem, ao integrar cada uma de suas aplicações, alavanca a melhor criação de complementos existentes que podem ser integrados. O *Streaming Service* e o *Cloud Computing* faz a comunicação ser intensa; a virtualização ser imediata, o armazenamento ser de grande amplitude e a estruturação e mapeamento de dados serem agilmente transferidos e fornecidos pela internet. (VELTE; VELTE; ELSENPETER, 2013).

5.5.4 Google

Fundada em 1998, seus idealizadores Larry Page e Sergey Brin com a visão de organizar toda a informação disponível e torná-las úteis. A Google implementa o mais satisfatório serviço visto na atualidade. Sua criação foi revolucionária, agregou várias plataformas em um desenvolvedor só, e priorizou produtos de apoio crucial a usuários e empresas que depende dos serviços de informação. (GOOGLE, 2016).

Com aplicação nas mais diversas áreas, a Google se legitima uma das melhores organizações no ambiente em que se estabelece. Cuidando de coordenar as diversas áreas de tecnologia da informação, ela trabalha

diariamente com dados apurados dos diversos meios possíveis. Acreditando que haverá utilidade de alguma forma, a Google sempre coleta dados de todos que a circundam e a acessam. Pensando em como será benéfico que quanto mais dado somar melhor para fomentação do seu sistema. (DAVENPORT, 2014).

A preocupação com os dados não se limita em apenas agregá-los, mas também em como serão reaproveitados, afinal informação útil é aquela que informa algo. Entra então o Big Data de forma para gerenciar seus arquivos. Ele não só auxilia como é o novo tratamento dessas informações e consegue definir mais uma forma de sistematizar os mecanismos de trabalho. Fazendo-o repensar e aprimorar a estrutura da sua organização e otimizá-la. (TAURION, 2015).

A Google possui e sempre se reinventa integrando diversos recursos em suas aplicações. A vantagem nítida de utiliza-la é tão óbvia que se tornou corriqueira. Atualmente não há um serviço de informação em que a Google não esteja envolvida, ou seja, inserida de alguma forma. Acaba se tornando uma mistura de serviços revolucionários com funcionalidades cotidianas. Suas invenções e renovações tecnológicas estendem em todos os âmbitos de dados, nada é rejeitado. (VELTE; VELTE; ELSENPETER, 2013).

As essências da tecnologia da Google neste primeiro momento têm-se concentrado na cloud computing. Seu servidor de email foi o primeiro a integralizar a ideia conjuntamente com a nuvem. Os seus serviços mesmo oferecidos individualmente, desenvolvem-se de forma aprimorada simultaneamente. Se existe uma atualização em uma aplicação, ela se estende em todos os âmbitos que lhe cabe. Não há uma só reforma onde não haja essencialmente uma mudança em outro recurso. E é dessa forma que a Google vem engrandecendo diante das melhorias tecnologias atuais. (VELTE; VELTE; ELSENPETER, 2013).

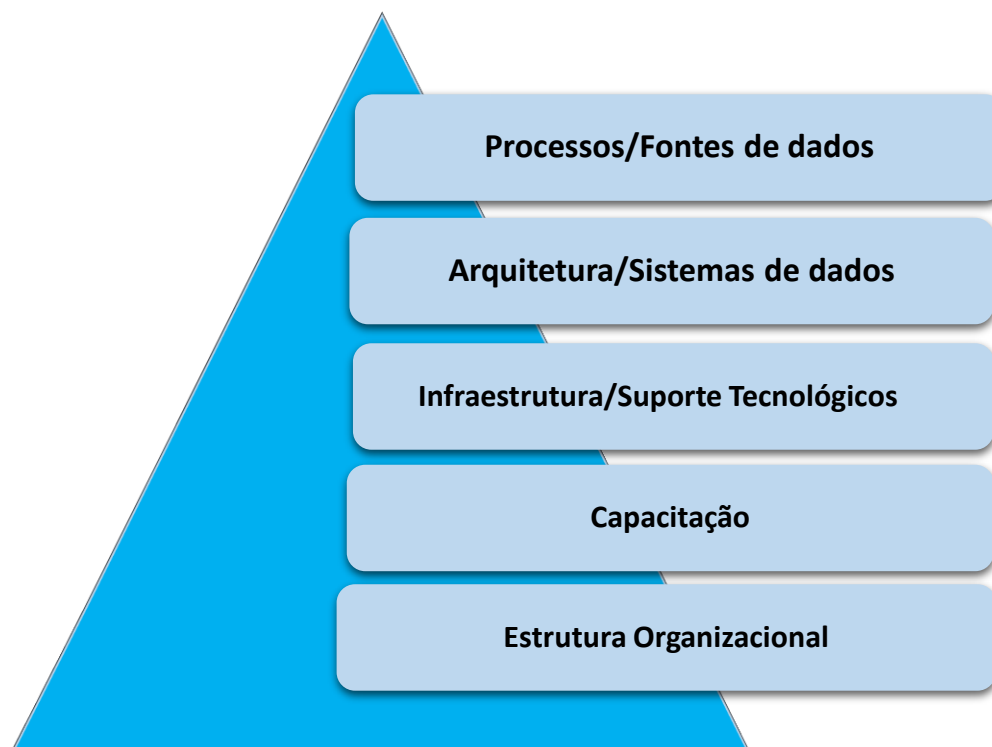
5.6 APLICAÇÕES EM BIBLIOTECAS

Com a tecnologia a favor de novos modelos para acessar a informação, têm-se trazido, cada vez mais diferentes meios de processamento de dados. Cabe aos centros de produção intelectual se adequar aos novos métodos de processamento de dados. Usá-los para que oriente os caminhos necessários para serem utilizados dentro de qualquer empresa, é o que tem tornado a análise dos serviços oferecidos interessantes ao provedor deles. O retorno que se obtém quando há uma nova necessidade ou que tal necessidade se encontra obsoleta é o que diferencia um bom serviço de outro, e uma empresa mediana de uma excelente. (TAURION, 2015; SORDI, 2015).

O impacto que o Big Data tem causado nas empresas de forma negativa ou positiva tem feito repensar toda a estrutura da informação (à forma de tratamento dos dados). De forma que a biblioteca, um polo de produção e resguardo de informação, não ficaria por fora de toda essa revolução de processamento de dados. E é a partir da ideia de como big data afetará a biblioteca remete-se a como as tecnologias serão adaptadas ou desenvolvidas por ela. (TAURION, 2015; SOUTO, 2010).

Ao implementar como funcionará o Big Data na biblioteca, primeiro precisa definir como e quais os processos que possam ser afetados. E como poderá ser feita essa implementação na biblioteca. Dentre as características possíveis que dialogam com o BD os aspectos voltados nas bibliotecas, possui âmbitos que atuam diretamente na transformação do BD, são elas: **fonte de dados, arquitetura e sistemas de dados, infraestruturas e suportes tecnológicos, capacitação e estrutura organizacional**. (TAURION, 2015; SORDI, 2015).

Figura 8 - Características nas Bibliotecas



Fonte: (TAURION, 2015)

5.6.1 Fonte de dados

Ao determinar o fator de dados, o primeiro passo é identificar o que são eles na instituição. Os dados existem na biblioteca e estão de forma explícita. Não há necessidade de cálculo para a sua obtenção, pois eles já existem e estão em evidência. Se uma biblioteca, por exemplo, levanta diariamente os números de quantos empréstimos são feitos, ou de quantas pessoas a frequentam, são exemplos de dados brutos que existem puramente antes mesmo da necessidade de serem analisados. (SORDI, 2015).

A origem do dado não é sempre o interessante. Não é se, não souber o que fazer com ele. O conjunto de dados se não administrados de forma que sua utilidade seja real para a instituição, continua sendo apenas um aglomerado de dados sem finalidade. Nesse ponto caberia responder quais dados são necessários para que se obtenha uma análise de sua influência dentro de um sistema de informação. (SORDI, 2015; TAURION, 2015).

Supondo que todos os dados existentes estão disponíveis, haverá uma seleção que dirá qual de fato é necessário para o sistema. E é nesse ponto em que se depara com a transformação de dados em informação. Para que a informação seja construída, é preciso que os dados sejam processados utilizando-se de análises que consigam extrair tais informações. (SORDI, 2015).

O processamento de dados é o método que, genericamente chamado assim, determinará a organização de toda a estrutura na instituição. Sem o devido tratamento ou a mediação humana para definir o seu desígnio, será puramente informações desconectadas e sem necessidade de se serem processadas. (SORDI, 2015).

Outras perguntas também necessárias relativas aos dados é se, as fontes são externas, ou apenas internas? São fontes confiáveis? São atuais? Uma vez que se necessita dos dados, é preciso determinar as características que são essenciais e tenham relevância. (TAURION, 2015; SORDI, 2015).

As fontes de dados estão ligadas a toda e qualquer tipo de dado que gerado sobre o sistema ou produtos do sistema. Um comentário, uma alusão ou divulgação, por exemplo, sobre um serviço ou produto valeria para ser avaliado de acordo. Contudo os moldes de tratamento de dados determinarão qual será a utilidade dos dados no sistema de informação. (SOUTO, 2010; TAURION, 2015).

Os dados internos são gerados dentro da empresa e podem servir de parâmetro para estipular e identificar quais dados pode ser utilizado. Será possível direcionar a sua utilidade, veracidade e o seu reuso, no sistema. Dados gerados por cada setor, ou que é possível contabilizar dentro da base de dados existente e, correlacionar essas informações. Históricos de clientes, com informações preditivas que possam auxiliar na mensuração das mesmas e serem úteis à empresa. (TAURION, 2015; DAVENPORT, 2014).

Já os dados externos, advêm da entrada de dados na empresa, de mídias exteriores acessadas. Toda uma informação fornecida externamente tem gerado outras informações que vão implementando os sistemas. Os dados, aqui chamados de dados transacionais, coletados pelas medidas de apuração externa com o ERP, por exemplo, tem demonstrado uma análise mais abrangente quanto aos históricos e armazenamento de acessos de algum

produto ou serviço, apontando assim, para a qualidade e amplitude do que foi oferecido. É nessa ótica que o usuário seria peça complexa de análise dos sistemas. Ele determinará como será e quais variáveis serão relacionadas ao sistema. (TAURION, 2015; SOUTO, 2010; DAVENPORT, 2014).

5.6.2 Arquitetura e Sistemas de dados

Com uma abrangência real e maior do que os modelos de gerência anteriores, o Big Data aparece como uma resolução de um grandioso modelo de dados. Ele vai explorando os aspectos externos dos dados e vai complementando a instituição de um jeito que aproveite ao máximo as informações reconhecidas nos dados. Com o direcionamento da análise voltado para todas as variáveis possíveis, e não como uma estimativa do que se tinha, a precisão com que os sistemas retornam as respostas transformou o Big Data em uma ferramenta preciosamente analítica. (TAURION, 2015; MAYER - SCHÖNBERGER; CUKIER, 2013).

Antigamente, usando-se a estimativa não se obtinha certeza de que os dados analisados realmente refletiam, e o que de fato ocorria com os sistemas. A ideia da estimativa dava a noção de uma amostra, para refletir em uma população inteira. Agora, com um sistema que consiga analisar todos os dados de forma integral e precisa, culminou em diferentes soluções apresentadas. Por meio da síntese de todos os parâmetros se tem a certeza de análises reais e íntegras e não apenas uma projeção do que realmente ocorria. Então a estimativa não se faz mais necessária. (TAURION, 2015; MAYER - SCHÖNBERGER; CUKIER, 2013).

Com o método analítico, consegue-se extrapolar as expectativas das análises e começa a ser possível a predição. Com os dados coletados, as transações efetuadas sobre certo produto ou serviço de uma empresa, fez com seja controlada todas as preferências do produto/serviço. A reutilização do dado também tem alterado as informações extraídas de algum contexto. Sendo possível controlar e modificar as quaisquer proposições existentes naquele momento. (TAURION, 2015; MAYER - SCHÖNBERGER; CUKIER, 2013).

5.6.3 Infraestruturas e Suporte tecnológicos

A infraestrutura está fundamenta em dois quesitos, na questão tecnológica e na execução das tecnologias utilizadas. Para administrar as escalas de volume, variedade e velocidade dos dados é concreta a funcionalidade de um software que consiga lidar com todos os aspectos ao mesmo tempo e continuamente. E ainda, como será a aplicação de determinada ferramenta utilizada. (TAURION, 2015; DAVENPORT, 2014).

No primeiro quesito é importante pensar que devido as grandes opções que existem atualmente, é preciso delimitar qual será a ferramenta que mais se adequará a necessidade do sistema. Cabe ao profissional que ao determinar a utilidade do software para otimizar o sistema, explore as carências existentes e preveja a potencialidade de tal aplicação. A estrutura tem que estar moldada de forma que haja a retroalimentação na qual garanta que o serviço seja avaliado e alterado, de acordo com a precisão e a comodidade do usuário. (SOUTO, 2015; TAURION, 2015).

O primeiro aspecto leva até o segundo que informará como se dará a aplicação do software. A partir da definição do programa, o desempenho fica comprometido pela forma com que ele será utilizado. É importante não só definir as características do sistema como também estipular os atributos que influenciarão sua produção. Como a organização trata os dados, como os armazena como os profissionais lidam com cada dado, como é a estrutura informacional e quais as os recursos existentes. Tudo isso influenciará na atuação do programa utilizado. (SOUTO, 2015; TAURION, 2015).

Considerando toda a tecnologia já existente e disponível, precisa-se levar em conta o suporte que mantém toda a estrutura preparada. É crucial que para o funcionamento de todas as demasiadas tecnologias na organização implantadas, possuam garantias que elas continuem funcionando. Fala-se na preparação do ambiente desde a instalação até assistência remota dos programas. Garantias que precisam ser consideradas quando se forem escolher os softwares em questão. (TAURION, 2015).

5.6.4 Capacitação

Os perfis dos profissionais têm mudado conforme a forma e a necessidade informacional se adaptam ao novo cenário de dilúvio de dados. Dentro da

perspectiva da capacitação, precisam-se observar os diferentes níveis que os envolve. Observar a questão do profissional, de suas competências e habilidades, das demandas práticas de sua aplicação. Entender a extensão de sua área na administração da informação, quais as técnicas e fundamentos, e a estrutura tecnológica que amplia e direciona o tratamento dos dados. (TAURION, 2015; SORDI, 2015; COSTA, 2014).

As atividades necessárias que precisam ser executadas surgem de como os profissionais se preparam para lidar com as informações e estruturas tecnológicas. Nesses quesitos, os profissionais da informação precisam se ater principalmente a questão da manipulação de dados. Da coleta até todo o processo que o engloba sintetizando e interpretando-o. Os bibliotecários são responsáveis em estruturar toda a informação disponível e administradas pelo software e por seus técnicos na instituição. Por serem responsáveis e conhecerem todo o conteúdo e a necessidade do usuário, serão a ponta final de toda a qualidade e integridade informacional. E é por isso que a competência dele está vinculada pela destreza com que ele lida com as tecnologias. (SORDI, 2015; COSTA, 2014).

Já os profissionais de tecnologia da Informação (TI), já naturalmente envolvidos com os âmbitos que as tecnologias trazem, são os colaboradores diretamente influenciadores das aplicações de TI. Sua responsabilidade aumenta na medida em que as tecnologias aprimoram. E com apoio deles que as áreas adjacentes, dão continuidade das atividades em suas funções. Esse profissional permite o seguimento do trabalho do bibliotecário, seja facilitando-o ou oferecendo o suporte. (SORDI, 2015).

Há também um novo tipo de profissional que além dos parâmetros de TI, dominam um pouco mais de analítica. O data scientist (cientista de dados) é uma nova tendência que ultrapassa a área de TI. Ele domina a ciência da computação com habilidades a mais para tratar das massas de dados. Ele é capaz de minerar os dados e extrair padrões, tendências e análises dos sistemas. Ainda consegue formular decisões estratégicas para que alavanquem as empresas e cuida da infraestrutura técnica necessária para dar suporte à organização. Uma especialização possível e extra de quem deseja ter um diferencial nas novas projeções existentes de tecnologia. (TAURION, 2015).

Assim, todas as áreas atualmente necessitam se atualizar. Principalmente as áreas que estão ligadas ao conhecimento e aprendizagem humana. A importância em assegurar o acesso à informação tanto às tecnologias é igualmente indispensável para a formação quanto para extensão do resultado da missão dos profissionais que a asseguram. (SORDI, 2015; COSTA, 2014).

5.6.5 Estrutura Organizacional

A estrutura organizacional é o conjunto de cada aspecto envolvido na dimensão de aplicabilidade do BD nas empresas e biblioteca. As questões de identificação dos dados, a arquitetura, infraestrutura e capacitação são os atributos que na proporção das suas características vão transformando o desempenho da organização. Nessa questão é preciso explorar todos os aspectos do Big Data. Parte-se então de cada item estudado e ampliado de forma que o utilize no sistema. (TAURION, 2015).

As proposições feitas serão avaliadas e colocadas em prática na demanda em que for requisitada. É aqui que se determinará se as análises feitas foram construtivas e os acertos necessários a se fazer. De alguma forma serão as respostas que foram exploradas por meio dos processos, extraído das análises feitas qualquer informação. Avaliando como os procedimentos foram feitos nesse momento se espera uma reação da empresa. O retorno obtido que dirá como a organização se comportará dentro de um futuro bem próximo, pois essa medida não pode ser esperada e é ajustada em tempo real, ou seja, muito rapidamente. (TAURION, 2015).

Também dentro desse parâmetro são definidas aqui as políticas necessárias que darão base para a logística do funcionamento do BD na organização. Desenhando as estratégias das tecnologias, de forma preditiva integrando os modelos de dados e desenvolvendo seu processamento. Os gestores são os responsáveis diretos das diretrizes estipuladas. As integridades das informações farão parte dos cuidados, o investimento imprescindível que deverá ser estabelecido, a questão da segurança e da privacidade da informação. Os relatórios que serão gerados das análises descritivas de cada função dentro da organização. Ou seja, estruturando todo o funcionamento e aplicação em sistema próprio. (TAURION, 2015).

5.6.6 A biblioteca, dados e informação

Levando em conta outras apurações relacionadas às bibliotecas, existem características que são abordadas mais tecnicamente e que podem influenciar cada processo internamente. Em suma, fala-se em como se dará a questão da extração de dados transformando-o em informação e, analisando todo o processo necessário para se obtê-la e convertê-la em conhecimento. Todo tratamento relevante à estrutura do conhecimento organizacional, será dado por como a entidade lidará com os ciclos informacionais e como administrará a informação. (SORDI, 2015).

É fundamental para a prática de gestão do conhecimento que se tenha zelo pela qualidade informacional. Essa dimensão só se pode ser alcançada por alguns aspectos que são relevantes que mantêm a integridade da informação. Presumindo que a informação é a interpretação e consolidação dos dados, os elementos fundamentados por essas características são classificados em: **atualidade, disponibilidade, confidencialidade e existência** da informação. (SORDI, 2015).

No fator atualidade é questionada a questão da efemeridade da informação. A temporalidade é ponto em que obriga os centros de documentação a sempre se alarmarem quanto a sua desatualização. Pois com o aumento do volume de dados e das informações nas organizações, esse processo pode ocorrer muito rápido e perder o valor informacional. (SORDI, 2015).

A disponibilidade esclarece a questão da acessibilidade à informação. Os recursos que podem ser disponibilizados, pelos sistemas de informação utilizando-se da infraestrutura tecnológicas, ou pelo tempo em que essas informações ficam à disposição do usuário. Nas bibliotecas físicas por exemplos, traz a prerrogativa do tempo juntamente com a acessibilidade onde o horário de funcionamento é modificador da variável da disponibilidade. (SORDI, 2015).

O aspecto confidencialidade faz alusão ao manuseio da confiança informacional. A relação de confiança do usuário com o sigilo garantido pelo sistema ou pelo anonimato creditado através da navegação via web faz desse atributo ser o mais alarmante. Parte do pressuposto que o indivíduo ou a entidade terão suas informações pessoais protegidas. Fala também sobre as

informações compartilhadas de forma voluntárias, mas sem perceber as possíveis ameaças. (SORDI, 2015).

E por fim, a existência da informação que abrange a condição da informação. Ela pode estar desestruturada ou estruturada, mas se encontra em acesso nos seus servidores. Nesse ponto explica-se a questão do armazenamento e do seu efetivo funcionamento se dedicando a guardar toda a informação absorvida ou criada pelos centros de documentação. (SORDI, 2015).

Para mensurar a qualidade, a manutenção, o armazenamento e a gestão de dados e as informações, tem como responsabilidade as entidades que cuidam e tratam de todo o tipo de informação colhida e de dado coletado. As competências necessárias de manuseamento de dado demandam prática nas aplicações nas bibliotecas. Sabendo como e quando utilizá-las mais preciso é confiável será seu processamento. (SORDI, 2015).

6 CONSIDERAÇÕES FINAIS

O presente trabalho objetivou apresentar as características primordiais do Big Data e suas aplicações. Foram listadas, em um primeiro segmento, as definições acerca do termo. Partindo-se de uma contextualização a ampla, foram elucidados e exemplificados os principais atributos, conforme sua estrutura. Desenvolveu-se, também, uma abordagem sobre o seu contexto histórico e as tecnologias envolvidas no BD.

Por conseguinte, o objetivo foi cumprido, pois por meio do levantamento bibliográfico foi possível esclarecer o tema, desenvolvendo as vertentes relacionadas com o Big Data. Dos aspectos investigados, a contextualização histórica favoreceu a sistematização das definições do termo, trazendo assim o real significado da sua representação e da sua tecnologia. Ao passo que, no desenrolar de suas tecnologias, foram averiguadas diferentes aplicações que foram fundamentais para a sintetização da sua ideia e vice-versa.

O estudo realizado trouxe uma compreensão das tecnologias e das aplicações do Big Data investidas nas empresas e bibliotecas. É visível que a evolução do Big Data para o caminho da aprendizagem é real e palpável, ao mesmo tempo em que a fomentação para sistemas de informação ainda se encontra principiando.

Como resultado deste trabalho, conclui-se que, após as expectativas trazidas pelo tema, o BD alcançará as mais diversificadas áreas. Visivelmente, dos levantamentos feitos, muito serviços já estão acontecendo e são ideais para as fundamentações futuras. São exemplos a *Cloud Computing* e o *Business Intelligence*, mantendo os sistemas informacionais ágeis e precisos.

A previsão para o potencial do Big Data ainda é incerta em decorrência da sua abrangência. É tangível a sua aplicabilidade em vários segmentos, principalmente aqueles que mais dependem da tecnologia. Mas a legítima preocupação é de que forma essa ferramenta será utilizada e por quem e com quais interesses. O controle existente é grande, e a sua potencialidade há um quê de exposição e análises corriqueiramente incertas referente ao contexto utilizado. Ainda há muito sobre o que e se preocupar.

Assim para um futuro estudo será necessário refletir as medidas realmente essenciais para a organização do conhecimento humano. Até onde chegará os recursos e aplicações de quantificação e qualificação da

informação humana. Até onde conseguirá separar a intenção informacional e a perda de privacidade. Se os métodos de gerenciamento informacional se darão sozinhos, de forma automática. As aplicabilidades nas áreas se serão possíveis, como serão e onde serão. E por fim, até onde será utilizado e por quanto tempo ainda será validada a capacidade humana.

REFERÊNCIAS

- ALMEIDA, José Maria Fernandes de. **Para uma História da Informática**. 1997. 21 f. Artigo - Curso de DSI - Sociedade da Informação, Departamento de Sistemas de Informação, Universidade do Minho, 1997. Disponível em: <http://repositorium.sdum.uminho.pt/bitstream/1822/875/1/HIS_TINF.PDF>. Acesso em: 05 de maio de 2016.
- ASHTON, Kevin. That 'Internet of Things' Thing: In the real world, things matter more than ideas. **RFID Journal**. Phoenix, p. 1-1. 22 jun. 2009. Disponível em: <<http://www.rfidjournal.com/articles/view?4986>>. Acesso em: 22 maio 2016.
- BATISTA, Emerson de Oliveira. **Sistema de Informação: o uso consciente da tecnologia para o gerenciamento**. São Paulo: Saraiva, 2004.
- CHEE, Brian J. S.; FRANKLIN JUNIOR, Curtis. **Computação em nuvem – Cloud Computing: Tecnologias e estratégias**. São Paulo: M.Books do Brasil, 2013. 247 p. Tradução de Mario Moro.
- COSTA, Maira Murrieta; CUNHA, Murilo Bastos da. O bibliotecário no tratamento de dados oriundos da e-science: considerações iniciais. **Perspect. ciênc. inf.**, Belo Horizonte, v. 19, n. 3, p. 189-206, set. 2014. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362014000300010&lng=pt&nrm=iso>. Acesso em : 26 junho 2016. <http://dx.doi.org/10.1590/1981-5344/1900>.
- COX, Michael; ELLSWORTH, David. **Application-Controlled Demand Paging for Out-of-Core Visualization**. NASA Ames Research Center. 1997. Disponível também em: <<http://www.nas.nasa.gov/assets/pdf/techreports/1997/nas-97-010.pdf>>. Acesso em: 22 de maio de 2016.
- DAVENPORT, Thomas. **Big data no trabalho: Derrubando mitos e descobrindo oportunidades**. Tradução: Cristina Yamagami. Rio de Janeiro: Elsevier, 2014.
- DEAN, Jeffrey; GHEMAWAT, Sanjay. MapReduce: Simplified Data Processing on Large Clusters. In: OSDI'04: SIXTH SYMPOSIUM ON OPERATING SYSTEM DESIGN AND IMPLEMENTATION, 6ª, 2004, San Francisco, Ca. **MapReduce: Simplified Data Processing on Large Clusters**. San Francisco: Google, 2004. p. 1 - 13. Disponível em: <<http://research.google.com/archive/mapreduce.html>>. Acesso em: 20 de julho de 2016.
- FACEBOOK. The Facebook produto e serviço. Disponível em: <https://www.facebook.com/FacebookBrasil/info?tab=page_info> Acesso em: 25 de junho de 2016.

FUNDABLE. **Amazon Startup Story**. 2017. Disponível em: <<https://www.fundable.com/learn/startup-stories/amazon>>. Acesso em: 16 de janeiro de 2017.

GARTNER; GLOSSARY, Gartner It Glossary. **Business Intelligence (BI)**. 2016. Disponível em: <<http://www.gartner.com/it-glossary/business-intelligence-bi/>>. Acesso em: 21 de junho de 2016.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002. 176 p.

GODOI, Lucas Moraes. História na educação matemática: uma proposta didática com o ábaco dos números inteiros. 2015. 69 f. TCC (Graduação) - Curso de Matemática, Instituto de Matemática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2015. Disponível em: <<http://www.lume.ufrgs.br/handle/10183/134206?show=full>>. Acesso em: 20 maio 2016.

GOOGLE. **Google Empresa**. Disponível em: <<https://www.google.com.br/about/company/>>. Acesso em: 25 de junho de 2016.

HADOOP, Apache. **What Is Apache Hadoop?** 2014. Disponível em: <<http://hadoop.apache.org/>>. Acesso em: 29 maio 2016.

HAHN, Seungrahn et al. **Capacity Planning for Business Intelligence Applications: Approaches and Methodologies**. 2. ed. Poughkeepsie, Ny: Redbooks, 2000. 214 p. Disponível em: <<http://www.redbooks.ibm.com/abstracts/sg245689.html?Open>>. Acesso em: 17 de julho de 2016.

IBM. **História**: Um pouco de história. Disponível em: <<http://www.ibm.com/br/ibm/history/>>. Acesso em: 15 maio 2016.

IVANOV, Todor; KORFIATIS, Nikolaos; ZICARI, Roberto V.. On the inequality of the 3V's of Big Data Architectural Paradigms: A case for heterogeneity. **Chair For Databases And Information Systems**: Institute for Informatics and Mathematics. Goethe University Frankfurt, 16 nov. 2013. p. 01-26. Disponível em: <<http://www.bigdata.uni-frankfurt.de/2013/11/inequality-3vs-big-data-architectural-paradigms-case-heterogeneity/>>. Acesso em: 29 de maio de 2016.

LANEY, Doug. **3D data management**: controlling data volume, velocity and variety. MetaGroup. 2001. Disponível também: <<http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>> Acesso em: 22 de maio de 2016.

LAURINDO, Fernando José Barbin; MESQUITA, Marco Aurélio de. Material Requirements Planning: 25 anos de história - Uma revisão do passado e prospecção do futuro. **Gest. Prod.**, São Carlos, v. 7, n. 3, p. 320-337, Dec. 2000. Disponível em:

<http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-530X20000003000009&lng=en&nrm=iso>. Acesso em: 15 de maio 2016.

LEE, In; LEE, Kyoochun. **The Internet of Things (IoT):** Applications, investments, and challenges for enterprises. *Business Horizons*, [s.l.], v. 58, n. 4, p.431-440, Jul. 2015. Elsevier BV. Disponível em:<<http://dx.doi.org/10.1016/j.bushor.2015.03.008>> Acesso em: 15 de maio 2016.

LIMA, Álvaro Vieira; LIMA, Davi Marques. **Business Intelligence como ferramenta gerencial no suporte ao processo de business performance management.** *Univ. Gestão e TI, Brasília*, v. 1, n. 1, p. 111-129, jan./jun. 2011.

MAYER SCHÖNBERGER, Viktor; CUKIER, Kenneth. **Big data:** como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana. Tradução: Paulo Polznoff Junior. Rio de Janeiro: Elsevier, 2013.163 p.

RASHID, Mohammad A.; HOSSAIN, Liaquat; PATRICK, Jon David. **The Evolution of ERP Systems:** A Historical Perspective. Idea Group Publishing. Chapter 1. In Enterprise Resource Planning: Global Opportunities and Challenges. 2002. Disponível em: <<https://faculty.biu.ac.il/~shnaidh/zooloo/nihul/evolution.pdf>> Acesso em: 19 de maio de 2016.

REINSCHMIDT, Joerg; FRANCOISE, Allison. **Business Intelligence Certification Guide.** California: Redbooks, 2000. 166 p.

SCHEPS, Swain. **Business Intelligence:** for dummies. Indianapolis: Wiley Publishing, Inc, 2008. 386 p.

SCUDERE, Leonardo. **Risco digital na web 3.0.** Rio de Janeiro: Elsevier, 2015. 154 p.

SHANNON, C. E.. A Mathematical Theory of Communication. **The Bell System Technical Journal**. S.l, p. 623-656. out. 1948. Disponível em: <<http://worrydream.com/refs/Shannon%20-%20A%20Mathematical%20Theory%20of%20Communication.pdf>>. Acesso em: 11 maio 2016.

SORDI, José Osvaldo de. **Administração da Informação:** fundamentos e práticas para uma nova gestão do conhecimento. 2. ed. São Paulo: Saraiva, 2015. 280 p.

SOUTO, Leonardo Fernandes. **Informação seletiva, mediação e tecnologia**: a evolução dos serviços de disseminação seletiva de informações. Rio de Janeiro: Interciência, 2010. 131 p.

TAURION, Cezar. **Big data**. Rio de Janeiro: Brasfort, 2015. 170 p.

VELTE, Anthony T.; VELTE, Toby J.; ELSENPETER, Robert. **Computação em nuvem**: uma abordagem prática. Rio de Janeiro: Alta Books, 2013. 335 p.

VERAS, Manoel. **Computação em nuvem**: Nova arquitetura de TI. Rio de Janeiro: Brasport, 2015. 175 p.

W3C. **A Little History of the World Wide Web**. Disponível em: <<https://www.w3.org/History.html>> Acesso em: 17 de maio de 2016.

W3C. **World Wide Web Consortium**. Disponível em: <<https://www.w3.org/>> Acesso em 17 de maio de 2016.

WINSHUTTLE. **Big Data and the History of Information Storage**. Disponível em: <<http://www.winshuttle.es/big-data-historia-cronologica/>>. Disponível também: <<http://www.winshuttle.com/big-data-timeline/>> Acesso em: 10 de maio de 2016.

WORLD WIDE WEB FOUNDATION. **History of the Web**. Disponível em: <<http://webfoundation.org/about/vision/history-of-the-web/>> Acesso em: 17 de maio de 2016.

ZIKOPOULOS, Paul et al. **Big Data Beyond the Hype**: A Guide to Conversations for Today's Data Center. [s.l]: Mc Graw Hill education, 2015. 393 p. Disponível em: <https://www-01.ibm.com/marketing/iwm/dre/signup?source=ibm-analytics&S_PKG=ov28197&dynform=11707&lang=en_US&S_TACT=000000VO&S_OFF_CD=10000733>. Acesso em: 20 de dezembro de 2016.